UNIVERSIDAD POLITÉCTICA DE MADRID

DOCTORAL THESIS

Aerodynamic database reconstruction via gappy high order singular value decomposition

Author:
Ana Isabel
MORENO LÓPEZ

Advisors:
Dr. José Manuel
VEGA DE PRADA

Dr. José Manuel
PERALES PERALES

March, 2016

Abstract

Aerodynamic database reconstruction via gappy high order singular value decomposition

by Ana Isabel MORENO LÓPEZ

A method based on an iterative application of high order singular value decomposition is derived for the reconstruction of missing data in multidimensional databases. The method is inspired by a seminal gappy reconstruction method for two-dimensional databases invented by Everson and Sirovich (1995) and improved by Beckers and Rixen (2003) and Venturi and Karniadakis (2004). In addition, the method is adapted to treat both noisy and structured-but-non-rectangular databases.

The method is calibrated and illustrated using a three-dimensional toy model database that is obtained by discretizing a transcendental function. The performance of the method is tested on three aerodynamic databases for the flow past a wing, one obtained by a semi-analytical method, and two resulting from computational fluid dynamics. The method is finally applied to an experimental database consisting in a non-exhaustive parameter space measurement of forces for a box-wing configuration.

Resumen

Reconstrucción de bases de datos aerodinámicas a través de la descomposición en valores sigulares para tensores de alto orden

por Ana Isabel MORENO LÓPEZ

Esta Tesis se centra en el desarrollo de un método para la reconstrucción de bases de datos experimentales incompletas de más de dos dimensiones. Como idea general, consiste en la aplicación iterativa de la descomposición en valores singulares de alto orden sobre la base de datos incompleta. Este nuevo método se inspira en el que ha servido de base para la reconstrucción de huecos en bases de datos bidimensionales inventado por Everson y Sirovich (1995) que a su vez, ha sido mejorado por Beckers y Rixen (2003) y simultáneamente por Venturi y Karniadakis (2004). Además, se ha previsto la adaptación de este nuevo método para tratar el posible ruido característico de bases de datos experimentales y a su vez, bases de datos estructuradas cuya información no forma un hiperrectángulo perfecto.

Se usará una base de datos tridimensional de muestra como modelo, obtenida a través de una función transcendental, para calibrar e ilustrar el método. A continuación se detalla un exhaustivo estudio del funcionamiento del método y sus variantes para distintas bases de datos aerodinámicas. En concreto, se usarán tres bases de datos tridimensionales que contienen la distribución de presiones sobre un ala. Una se ha generado a través de un método semi-analítico con la intención de estudiar distintos tipos de discretizaciones espaciales. El resto resultan de dos modelos numéricos calculados en CFD. Por último, el método se aplica a una base de datos experimental de más de tres dimensiones que contiene la medida de fuerzas de una configuración ala de Prandtl obtenida de una campaña de ensayos en túnel de viento, donde se estudiaba un amplio espacio de parámetros geométricos de la configuración que como resultado ha generado una base de datos donde la información está dispersa.

Contents

List of Abbreviations

BR	**B**eckers and **R**ixen
C_D	Drag Coefficient
CF	**C**onvergence **F**actor
CFD	**C**omputational **F**luid **D**ynamics
C_L	Lift Coefficient
EB	**E**rror **B**ound
ES	**E**verson and **S**irovich
HOSVD	**H**igh **O**rder **S**ingular **V**alue **D**ecomposition
L.E.	**L**eading **E**dge
MaxE	**Max**imum **E**rror
MSE	**M**ode **S**election **E**stimate
POD	**P**roper **O**rthogonal **D**ecomposition
RMSE	**R**oot **M**ean **S**quare **E**rror
SVD	**S**ingular **V**alue **D**ecomposition
UAV	**U**nmmaned **A**erial **V**ehicle
VLM	**V**ortex **L**attice **E**ethod
VK	**V**enturi and **K**arniadakis

Chapter 1

Introduction

In the field of aerospace industry, multidimensional databases are daily managed for many purposes related to, e.g., design, certification, flight control and flight simulators. They are generated as a result of collecting data from a big amount of studies realised by several techniques of analysis that in most of cases represent different level of accuracy in the results of the same physical problem. The various dimensions in these databases are usually associated with either spatial and temporal coordinates and/or parameters. In particular, aerodynamics is especially known for generating huge databases and for integrating different tools which costs and accuracy are very diverse. For instance, both computational fluid dynamics (CFD) simulations and wind tunnel campaigns produce big amount of data but their cost and restriction to time schedule can reduce the number of tested and calculated cases generating sparse databases. Also, note that databases generated from experiments include errors, due to imprecisions and other practical problems, i.e. broken sensors.

Databases containing aerodynamic data are *strongly correlated* both in the physical space, time dimensions and in the parameter space, due to *redundancies* produced by physical laws, such as those associated with the underlying Navier-Stokes equations. Because of these redundancies, the *actual information* contained in the database is much smaller than the *database size*, which could become huge due to the curse of dimensionality [1] as the number of database dimensions increases. This Thesis is directly focused on treating incomplete, noisy and strongly correlated data making aerodynamic databases good candidates to be studied here, but the proposed methods can be easily extrapolated to many other fields (usually less demanding).

Moreover it is well known that the problem of the existing technology is related nowadays with the processing of information rather than collecting it. Due to this, it is really important to pre-process the data, increasing the quality of databases and reducing its size to be efficiently managed. This challenge involves various tasks, such as database compression, reconstruction of lost data and error filtering, aspects that will be analysed in this work.

Databases described in multiple dimensions can be represented as higher order tensors. Intuitively, tensors are the multi-linear generalization of matrices

(tensors of order two) or vectors (tensors of order one), since they can be considered as a set of numbers organized in three or more directions. Identifying the actual information along such directions is the main ingredient to repair and/or complete the database and therefore, this is the main scope of the *multi-way data analysis* [2]: it is the extension of two-way data analysis to higher order datasets. Since the beginning of the 1960s (Tucker's work [3, 4] laid the groundwork for this methodology), a series of techniques have been devised specifically aimed at treating with three-way data. These techniques were later extended to higher-way data [5] representing a milestone in multi-way literature linking multilinear algebra with the models originated in psychometrics and chemometrics.

First approach to deal with multidimensional databases is to consider them as families of vectors, isolating one of the dimensions and folding the remaining dimensions together into a single dimension; this means that the number of dimensions can be always reduced to two. For instance, three-dimensional databases resulting from time-dependent 2D flows can be treated as collections of vectors, each giving a snapshot at a particular value of time (see, e.g., [6]). Truncated *proper orthogonal decomposition* (POD), which is a variant of *singular value decomposition* (SVD), takes advantage of the redundancies among the vectors and produces a lower dimensional approximation. For example, transient three dimensional flow can be analysed by performing POD analysis at time- and space- windows of the flow in order to separately study temporal phenomena or focus the analysis on sub-domains [7]. This technique is useful in a variety of contexts, such as database generation [8, 9, 10], derivation of reduced order models [11, 12, 13, 14, 15, 16], compression [17, 18], and image processing [19].

However, the mentioned two-dimensional reduction has an important drawback: it looses the individual redundancies along the dimensions that have been folded together. Instead, it is possible to account for such redundancies using *high order singular value decomposition* (HOSVD), that will be thoroughly used along in this work. This is an extension to tensors [20, 21, 22] of standard SVD [23], a well known mathematical decomposition that has largely demonstrated its usefulness in many fields. Unfortunately, the SVD only applies to matrices. The use of HOSVD overcomes this limitation and the advantage will be clearly appreciated in both database generation [24] and compression [25]. For example, in the latter case, the compression factor increases exponentially with the number of the database dimensions whose redundancies are accounted for. Thus, identifying as many redundancies as possible in the database strongly alleviates the curse of dimensionality. The advantages of the HOSVD description of a multidimensional database are widely known as its application to very different fields demonstrates: signal processing [26], data clustering [27] and others. Concerning to aerospace industry applications and following a former idea [8, 9] in the context of standard POD (for two-dimensional databases), HOSVD can be combined with interpolation [25] to calculate the database elements at intermediate values of the physical quantities accounted for in the

various dimensions. Such combination has been successfully used to speed up calculations in both the real time control of reciprocating engines [28] and the conceptual design of aircraft components [29].

As mentioned, one of the most important tasks in multidimensional databases is filling missing data where this Thesis is mainly focused. This problem can be addressed from different approaches, for example one of them is HOSVD combined with interpolation that has been demonstrated to be very efficient since HOSVD allows for reducing multi-dimensional interpolation to a series of one-dimensional interpolations. However, it is not extensively applicable to all gappy reconstruction problems and the accuracy of the results are prone to scheme interpolation errors. *Gappy reconstruction*, consisting in reconstructing missing data, is needed in, e.g., experimental wind tunnel databases due to a complete obstruction of sensors or the impossibility of locating sensors in a part of the physical domain. The theoretical as well as economical importance of solving these problems is straightforward. This task has been addressed in the literature in various contexts, since the pioneering work by Yates [30], who applied least squares to fill in missing data.

Gappy reconstruction methods can be either statistical [31] or deterministic [6, 8, 9, 32, 33, 34, 35]. Several statistical methods lead with missing values but among them, Kriging-based methods are the most commonly applied. Unlike other estimation procedures, Kriging provides a measure of the error and associated confidence intervals [36]. Each approach works better or worse depending on some assumptions and certain characteristic of the database construction. For instance, a detailed comparison between POD-based methods and Kriging interpolation [6] concludes that the latter is more effective when large gappyness or black zones are present. Concentrating on deterministic approaches, they aim at reconstructing one or more gappy vectors by generating a decomposition of the database into a series of data-based orthogonal functions. The Karhunen-Loeve (KL) expansion, also known as POD, is an optimal decomposition in terms of minimizing the mean square error between the database and its truncated reconstruction and also minimizes the number of orthogonal functions needed to describe the database for a given error [37]. This is the cause/consequence of the mentioned property of POD: being able to take advantage of the database redundancies, generating a suitable lower dimensional approximation. Regarding the knowledge about the database, it may happen that:

- A low dimensional approximation of the exact database is known beforehand. The gappy vectors are reconstructed one-by-one, imposing that they be close to the low dimensional approximation.
- Non a priori knowledge about the database is available.

The second situation is obviously the most difficult one since we do not have a starting point. Everson and Sirovich [32] (hereafter ES) dealt with both situations and they succeeded in developing a two-step method, by:

(i) First, reconstructing some of the gappy vectors and calculating a low dimensional approximation of the database. To perform this step, they developed a genuine gappy reconstruction method.

(ii) Then, using the low dimensional approximation calculated in step 1 to reconstruct any new gappy vector. The method to perform this step could be seen as an inverse design method.

The ES method has provided the basis for many applications and derived methods. In particular, Bui-Tanh et al.[9] considered this method to perform steps (i) and (ii) as independent methods, and applied them to some databases resulting from the inviscid aerodynamic flow around an airfoil.

Here, we use part of the ES *gappy-reconstruction method*, in particular the step (i), which is based on an iterative algorithm. At each iteration step, POD is applied to the database in which the result of the former step is used as a new guess for gappy data, while the original data are maintained at non-gappy positions. This method was seminal, but exhibits two great difficulties: the final error strongly depends on both the initial guess and the number of retained POD-modes. As mentioned, if we are dealing with the second situation, we have non a priori knowledge on how to choose the initial guess. Successively, Beckers and Rixen [33] and independently, Venturi and Karniadakis [34], modified this step introducing a second iteration on the number of retained modes. This solution solved both the mentioned difficulties, yielding fairly robust approximations in cases in which the standard ES gappy-reconstruction method fails, rendering the original algorithm stronger in case of using a bad initial guess. The modified method will be referred to as the ES-BR-VK method. The improved method has been extensively used in several applications [38, 39, 40, 41]. For instance, Gunes et al.[6] have extensively tested it in families of flow snapshots calculated using a CFD solver for specific values of time in the 2D unsteady flow around a cylinder. As previously mentioned, this work performs an exhaustive comparison of the ES-BR-VK method with Kriging interpolation. We recall that the POD-based method generally works better than Kriging for small gappyness/large temporal resolution, while Kriging is preferred when either the gappyness level is high or the temporal resolution is small. In particular, the ES-BR-VK method does not produce any results (and Kriging interpolation was proposed instead) when gappy data include a fixed spatial region (*black box*) for all values of time, which is needed to reconstruct experimental data in spatial regions that are not accessible to measurements.

As formulated, the ES-BR-VK method applies to (one-dimensional) families of vectors, which can also be seen as matrices. The first goal of the present Thesis is to extend the method to multidimensional databases using HOSVD. The resulting method (which will be referred to as *gappy-HOSVD* method) works fairly well, both when gappy elements are randomly located along the database and when they are concentrated, and also when gappyness is located black-box-like. Both are common characteristics of industrial databases which have not been always successfully addressed with same tools, as mentioned above.

In addition, the method will also be adapted to treat noisy databases, which are frequent in industry and, as already noticed in [34], produce a non-monotone behaviour of the ES-BR-VK method. In other words, the performance of the gappy reconstruction method worsens as the number of retained modes increases beyond some points. This is a well known problem, which requires to select the appropriate number of modes in order to leave the noisy errors out. This is a not easy task, and often it is addressed manually when the database size is small enough or a priori knowledge of the data is known. Usually this is not the case of aerospace databases. In this Thesis, a method will be presented that is synergic with HOSVD and dynamically selects well the 'optimal' numbers of modes along the various database dimensions, without the human intervention. The resulting method will be referred to as the *gappy-noisy-HOSVD* method.

On the other hand, the standard application of POD, SVD, and HOSVD can only be made to structured, rectangular/hyper-rectangular databases. Unfortunately, these databases may either be not possible or not convenient because some of the database positions needed to complete the hyper-rectangle are either non-physical (as in, e.g., the flow around an obstacle when Cartesian coordinates are used) or of no interest (e.g., some combinations of the Mach number and the angle of attack in the flow around an aircraft wing). The idea we use in order to obtain an HOSVD description in a non-rectangular mesh is to apply the gappy-HOSVD method in an augmented hyper-rectangular database considering as gappy elements those that have been added to the original (non-rectangular) database. In other words, the method will fill in the spurious added data in such a way that they are consistent (along the various dimensions) with the original database elements.

Summarizing, the scope of this Thesis is two-fold, since it shall (i) extend the ES-BR-VK method to treat multidimensional databases and (ii) adapt the method to cope with noisy and non-rectangular databases. We note that efficient 'universal methods' to repair/complete industrial databases are not to be expected. This is because these tasks most likely depend both on the structure of the underlying redundancies and on the magnitude of background errors (e.g., CFD-generated errors), which are unavoidable in practice. Thus, any method (in particular, the methods developed here) that is intended to work reasonably well may involve a few parameters that should be manually calibrated for each specific application (e.g., aerodynamics). Even so, we remark the fact that the methods developed in this Thesis will be robust, namely the sensibility of the results on the tunable parameters will be small. Also, we note that no assumptions will be made on the nature of the redundancies the methods are based upon. The only requirement is that such redundancies be present. This generality is inherited from the original ES method and makes a difference from gappy reconstruction/interpolation methods based on, e.g., polynomial or radial basis functions, and also with traditional error filtering methods that assume specific error properties, such as Gaussian distributions or zero means. These assumptions strongly limit the scope of these traditional methods. Here, instead, the

gappy-noisy-HOSVD method does not consider any assumptions about noise nature. This characteristic set the bases for a new filtering method [42]. Furthermore, taking advantage of the well known property of HOSVD-based methods in identifying patterns and consequently constructing lower dimensional models, the methods developed in this Thesis can be further improved developing a new method able to efficiently determine critical positions in a multidimensional database which are best candidates to be tested in order to improve the fidelity of the reconstructed database. This concept has been already addressed in literature, mainly based on POD methods [43, 44].

The methods developed in this Thesis will be first tested using a *toy model database*, which is obtained discretizing a transcendental function (see Figure 2.2). This database is 'clean', which helps to separate difficulties, allowing for both identifying the main requirements of the methods and testing their performance. The methods will also be applied to three aerodynamic databases containing the pressure distribution around a wing for various values of angle of attack. The first database is obtained by a semi-empirical method and the remaining two by using CFD, which inherently contains some level of noise. All applications will be performed using MATLAB on a desktop PC, with a i7-4930K CPU @ 3.40GHz processor.

The remaining of the work is organized as follows. The basic ability of POD/SVD and HOSVD to compress databases and filtering errors out will be recalled in §2, where notation will be established. In addition, the above mentioned toy model database is described in the same chapter, §2.3. The iterative ES method and the doubly iterative BR-VK extension will be revisited in §3, where the extended gappy-HOSVD and gappy-noisy-HOSVD methods for multidimensional databases mentioned above will be developed. The toy model database applications are included in §3 in order to better illustrate the methods performance. These methods will be applied in §4 and §5 to the aerodynamic databases mentioned above. Furthermore, §6 presents the results obtained by the proposed methods application to an experimental database that includes the aerodynamic coefficients of a box-wing configuration. This database comes from a wind tunnel test campaign which scope was studying the influence of several geometric parameters in those coefficients, by varying them from a baseline configuration model. The Thesis ends with some concluding remarks, in section §7, and highlighting the open research lines as future work.

Chapter 2

Mathematical Background

This chapter recalls the mathematical background on which the techniques developed in this Thesis are based. More than giving a complete description of this mathematical basis, the intention is pointing out main concepts needed to understand next chapters. Main features of these techniques will be illustrated with a toy model defined in section §2.3. Furthermore, the nomenclature and terminology used all along this work is mainly defined in this chapter.

2.1 POD and SVD

The POD of a system of vectors and the SVD of the associated matrix (whose columns are the vectors) are concepts closely related to each other, even though they have a different origin. SVD seems to have been first invented [45] independently by Beltrami and Jordan (in 1873-74) for square matrices associated with bilinear forms, and extended for rectangular matrices by Eckart and Young [46]. The SVD of a $(I \times J)$-matrix A is given by [23]

$$A = U\Sigma V^\top \quad \text{or} \quad A_{ij} = \sum_{l=1}^{r} \sigma_l U_i^l V_j^l, \tag{2.1}$$

where r is the rank of A. Here, the columns of the $I \times r$-matrix U (U_i^l refers to the column vector l of i elements) and the $J \times r$-matrix V (V_j^l refers to the column vector l of j elements), known as the *left* and *right SVD-modes*, respectively, are the (orthonormal) eigenvectors associated with strictly positive eigenvalues of the positive semi-definite matrices $AA^\top$ and $A^\top A$, respectively. The elements of the diagonal $r \times r$-matrix Σ, $\sigma_1, \ldots, \sigma_r$, known as the *singular values*, are the square roots of the associated eigenvalues. However, round-off errors have a dramatic effect if computations are performed using the spectral decomposition of the matrix $AA^\top$ or $A^\top A$. Instead, a more efficient method [23] is implemented in the MATLAB command 'svd' that solves this difficulty. The theoretical procedure is improved by applying a QR decomposition to the matrix A, as $A = Q \cdot R \cdot E$, where Q is orthogonal, R is upper triangular and E is a column permutation matrix. Thus, $A^T A = E^T \cdot R^T \cdot R \cdot E$, which is quite advantageous when the number of rows of A is much larger than the number of columns.

Once the right SVD-modes have been calculated as the eigenvectors of $A^T A$ associated with non zero singular values, the left SVD-modes are calculated as $U^k = \sigma_k^{-1} A \cdot V^k$. If the number of columns is larger than the number of rows, this concept is applied to A^T instead.

Now, if the singular values are sorted in decreasing order and Eq. (2.1) is truncated to $m < r$ terms, the root mean square (RMS) truncation error is

$$\text{RMS} = \sqrt{\frac{1}{IJ} \sum_{i,j} \left(A_{ij} - \sum_{l=1}^{m} \sigma_l U_i^l V_j^l \right)^2} = \sqrt{\frac{1}{IJ} \sum_{l=m+1}^{r} \sigma_l^2}. \tag{2.2}$$

Moreover, for a given $m < r$, SVD-modes and associated singular values minimize the Eq. (2.2), meaning that SVD is an *optimal decomposition*. Note that truncated SVD yields a very good approximation of the matrix when the neglected singular values are small, which occurs when appropriate redundancies occur among the rows and columns of the matrix A. In other words, truncation means retaining only the most energetic (measuring energy with the RMS norm) modes, and the neglected modes are orthogonal to the retained modes, which means that highly oscillatory, noisy errors are filtered out by truncation. In general, such automatic filtering effect depends on both the magnitude and structure of errors.

The POD of a system of J vectors, $\mathbf{u}_1, \ldots, \mathbf{u}_J \in \mathbb{R}^I$, is a concept that seems to have been first obtained by Pearson [47] in 1901. The aim is to calculate a new system of orthonormal vectors, $\mathbf{U}_1, \ldots, \mathbf{U}_J$, known as *POD-modes*, such that, for each $n \leq r =$ rank of the system of vectors, the span of $\mathbf{U}_1, \ldots, \mathbf{U}_n$ is the vector subspace of dimension n that is closer to the system $\mathbf{u}_1, \ldots, \mathbf{u}_J \in \mathbb{R}^I$. It follows that $\mathbf{U}_j = \sum_{k=1}^{J} \alpha_j^k \mathbf{u}_k$, where the coefficients α_j^k are given by the eigenvalue problem [48]

$$\sum_l R_{kl} \alpha_j^l = \sigma_j^2 \alpha_j^k, \text{ with } R_{ij} = \langle \mathbf{u}_i, \mathbf{u}_j \rangle, \tag{2.3}$$

The scalars $\sigma_j \geq 0$ are known as the *POD singular values* and the positive semi-definite matrix R, as the *covariance matrix*. The *inner product* $\langle \cdot, \cdot \rangle$ can be customized. Calculating POD through the covariance matrix is quite sensitive to round-off errors, which is avoided in a similar fashion as mentioned above in connection with SVD. In fact, using the Euclidean inner product, the POD-modes and singular values coincide with the left SVD-modes and singular values of the matrix A whose columns are the vectors $\mathbf{u}_1, \ldots, \mathbf{u}_J$.

2.2 HOSVD

The natural extension of SVD to third order tensors (*cf.* Eq. (2.1)) is $A_{ijk} = \sum_{l=1}^{r} \sigma_l U_i^l V_j^l W_k^l$, which is known as the *canonical decomposition* [22]. The minimum value of r such that this decomposition exists is called the *rank of the tensor*. Unfortunately, determining the rank for general, larger than two, order tensors is an open problem [22]. Furthermore, the computation of the associated minimal decompositions involves an optimization problem that is quite computationally expensive and generally ill-posed [22, 49].

For the sake of clarity, we consider third-order tensors, but higher order tensors are treated similarly.

Thus, some more flexible decompositions, known as Tucker decompositions [22], are considered in which σ depends on three indexes, namely

$$A_{ijk} = \sum_{i'=1}^{r_1} \sum_{j'=1}^{r_2} \sum_{k'=1}^{r_3} \sigma_{i'j'k'} U_i^{i'} V_j^{j'} W_k^{k'}. \tag{2.4}$$

This decomposition can be performed in various ways, depending on how the *core tensor* σ and the three sets of *modes*, $\{U_i^1, \ldots, U_i^{r_1}\}$, $\{V_j^1, \ldots, V_j^{r_2}\}$, and $\{W_k^1, \ldots, W_k^{r_3}\}$, are defined. In 1966, Tucker [4] set the preliminar definition of a decomposition that has been more recently popularized by de Lathauwer et al. [20] as Higher Order SVD (HOSVD), a convincing generalization of SVD to tensor decomposition, in which modes along the various directions of the tensor are the POD-modes of the associated tensor *fibers* (namely, the vectors that are obtained fixing two of the indexes of the tensor, see Figure 2.1); noting that the HOSVD-modes are orthonormal, the core tensor is given by

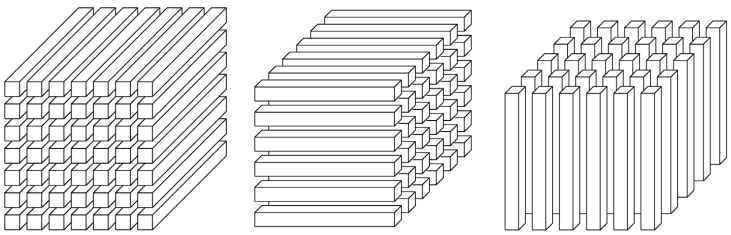

FIGURE 2.1: The fibers of a third order tensor.

$$\sigma_{i'j'k'} = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} A_{ijk} U_i^{i'} V_j^{j'} W_k^{k'}. \tag{2.5}$$

HOSVD is not optimal, but it is nearly optimal since (i) HOSVD-modes are POD-modes of the fibers of the tensor and thus they provide best RMS approximations of fibers; and (ii) once the modes have been calculated, the orthogonal projection implicit in (2.5) means that the core tensor optimizes the projection in the sense of mean squares.

The core tensor and the HOSVD-modes are very efficiently calculated using the MATLAB function 'hosvd' (included in the MATLAB toolbox TP Tool), which is able to treat tensors of any order. The calculation is performed in two steps: first, the three sets of HOSVD-modes in (2.4), and secondly, the core tensor as in Eq. (2.5), resulting from projecting the whole tensor onto the set of tensor products of the HOSVD-modes. Since first step involves a standard POD, computational efficiency is greatly improved applying a QR decomposition, as explained in §2.1. The calculation of HOSVD-modes and the core tensor are both mathematically well posed and computationally inexpensive (compared with other decompositions), which is specially relevant when the decomposition is to be applied a large number of times, as it will happen in the iterative methods presented in §3.

The HOSVD in Eq. (2.4) is exact if r_1, r_2, and r_3 are the ranks of the fibers along the directions i, j, and k, respectively. On the other hand, if the elements of A show redundancies along the three dimensions, *truncating the expansion* in Eq. (2.4) to $m \leq r_1$, $p \leq r_2$, and $q \leq r_3$ terms, as

$$A_{ijk} \simeq A^{\text{trunc}}_{ijk} \equiv \sum_{i'=1}^{m} \sum_{j'=1}^{p} \sum_{k'=1}^{q} \sigma_{i'j'k'} U_i^{i'} V_j^{j'} W_k^{k'}, \tag{2.6}$$

may still give a good approximation. In fact, based on the root mean square (RMS) norm definition of a tensor X

$$||X||_{\text{RMS}} = \sqrt{\frac{1}{IJK} \sum_{i,j,k} X_{ijk}^2}, \tag{2.7}$$

the RMS *truncation error* can be estimated/bounded in terms of an a priori *error bound* (EB), as

$$||A - A^{\text{trunc}}||_{\text{RMS}} \leq \text{EB} \equiv \sqrt{\frac{1}{IJK} \left(\sum_{i'=m+1}^{r_1} \alpha_{i'}^2 + \sum_{j'=p+1}^{r_2} \beta_{j'}^2 + \sum_{k'=q+1}^{r_3} \gamma_{k'}^2 \right)}, \tag{2.8}$$

where $\alpha_{i'}$, $\beta_{j'}$, and $\gamma_{k'}$ are the HOSVD singular values of the fibers of the tensor along the i, j, and k directions, respectively. Note that EB is an upper bound, while its counterpart for the standard SVD, in Eq. (2.2), is exact. Now, the numbers of modes retained in truncated HOSVD along the various directions of the tensor will be selected below to minimize the error bound EB defined in Eq. (2.8), which is easily performed.

On the other hand, HOSVD is a genuine extension to tensors of standard SVD because the 2D restriction of Eq. (2.4), namely $A_{ij} = \sum_{i',j'=1}^{r} \sigma_{i'j'} U_i^{i'} V_j^{j'}$, is such that the matrix $\sigma_{i'j'}$ is diagonal; in higher dimensions, instead, the core tensor is not generally diagonal.

The third order tensor can also be treated via standard SVD, isolating one of the indexes (say, the first index). The associated SVD is

$$A_{ijk} = \sum_{l=1}^{r_1} \alpha_l U_i^l V_{jk}^l, \tag{2.9}$$

where the left modes $\{U_i^1, \ldots, U_i^{r_1}\}$ and the singular values $\alpha_1, \ldots, \alpha_{r_1}$ are still given by the POD of the fibers along the index i, while the right modes $\{V_{jk}^1, \ldots, V_{jk}^{r_1}\}$ are given by $V_{jk}^l = \alpha_l^{-1} \sum_i A_{ijk} U_i^l$. But this decomposition only takes into account redundancies in the first index i and thus yields a less efficient truncation than HOSVD: truncating to $m < r_1$ terms in Eq. (2.9), the required storage capacity is $m + m(I + JK)$, which (assuming that $m \ll I$, $p \ll J$, $q \ll K$, and $I \sim J \sim K$) scales as mJK. Instead, truncation to (m, p, q)-terms in the HOSVD (see Eq. (2.6)) requires a storage capacity of $mpq + mI + pJ + qK$, which is much smaller than mJK. In fact, taking into account that HOSVD is orthogonal, the required storage capacity, which will be referred to as *truncated information* (TI), is

$$\text{TI} = mpq + \frac{(2I - m - 1)m + (2J - p - 1)p + (2K - q - 1)q}{2}. \tag{2.10}$$

Thus, HOSVD is much more efficient than standard SVD in multidimensional databases to identify the relevant information. This is important because identifying the relevant information is the main ingredient in database compression, reconstruction, completion, and error filtering.

2.3 A toy model database

The mathematical methods presented in the next chapter will be illustrated using a 3D database A (namely, toy model) defined as follows. It is a third order tensor that results from a linear transformation (Eq. (2.12)) applied to the transcendental function f, shown in Eq. (2.11),

$$\begin{aligned} f(x, y, z) = {} & x^2 \left[\sin\left(5\pi y + 3\log(x^3 + y^2 + z + \pi^2)\right) - 1\right]^2 - 4x^2 y^3 (1 - z)^{3/2} \\ & + (x + z - 1)(2y - z)\cos\left(30(x + z)\right)\log(6 + x^2 y^2 + z^3) \end{aligned} \tag{2.11}$$

$$T : \mathbb{R} \to \mathbb{R} \quad T(f(x, y, z)) = 2 \cdot (f(x, y, z) - f_{min})/(f_{max} - f_{min}) - 1 \tag{2.12}$$

such that the resulting function varies between -1 and 1 in the cubic domain $0 \leq x, y, z \leq 1$. f_{max} and f_{min} are the maximum and the minimum values that the function f adopts in such domain. This transformation facilitates the interpretation of errors. Denoting the transformed function also as f, the tensor

A is defined as

$$A_{ijk} = f(x_i, y_j, z_k), \quad \text{with } x_i = \frac{i-1}{I-1}, y_j = \frac{j-1}{J-1}, z_k = \frac{k-1}{K-1}, \tag{2.13}$$

for $i = 1, \ldots, I$, $j = 1, \ldots, J$, and $k = 1, \ldots, K$.

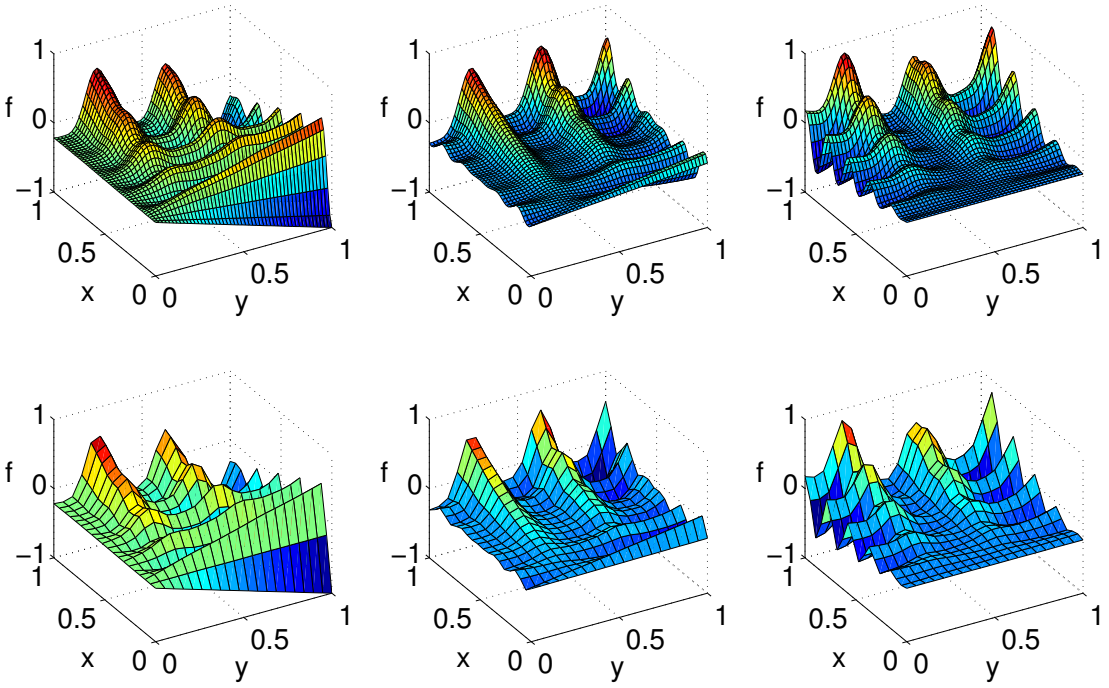

FIGURE 2.2: The dense (top) and coarse (bottom) toy model databases represented at constant z-value sections: $z = 0$ (left), $z = 1/2$ (middle) and $z = 1$ (right).

Figure 2.2 shows the sections $z = 0$, $z = 1/2$ and $z = 1$ of the discretized function, whereas Figure 2.3 shows the sections $x = 0$, $x = 1/2$ and $x = 1$, with $I = J = K = 50$ (top) and $I = J = K = 20$ (bottom). These will be referred to as the *dense* (with $50 \times 50 \times 50 = 125,000$ data points) and *coarse* (with $20 \times 20 \times 20 = 8,000$ data points) *toy model databases*, respectively. As can be seen, f is oscillatory in the three directions and does not exhibit any simple regularity, except for being defined through sums, products, and functional composition of the sine, cosine, fractional power and log functions. In particular, that term proportional to $(1 - z)^{3/2}$ prevents the function from being analytic. Note that the function exhibits different patterns in different locations. The 20^3 mesh is quite coarse, especially in the x direction, where the function exhibits five oscillations, which means that the discretized database contains only four points per period. All these make the toy model database quite demanding for any method (in fact, much more demanding than the usual databases in practical situations), except of course for the fact that the toy model database is 'exact' in the sense that it is described by an algebraic expression, free from errors.

For illustration, the three sets of HOSVD singular values associated with the

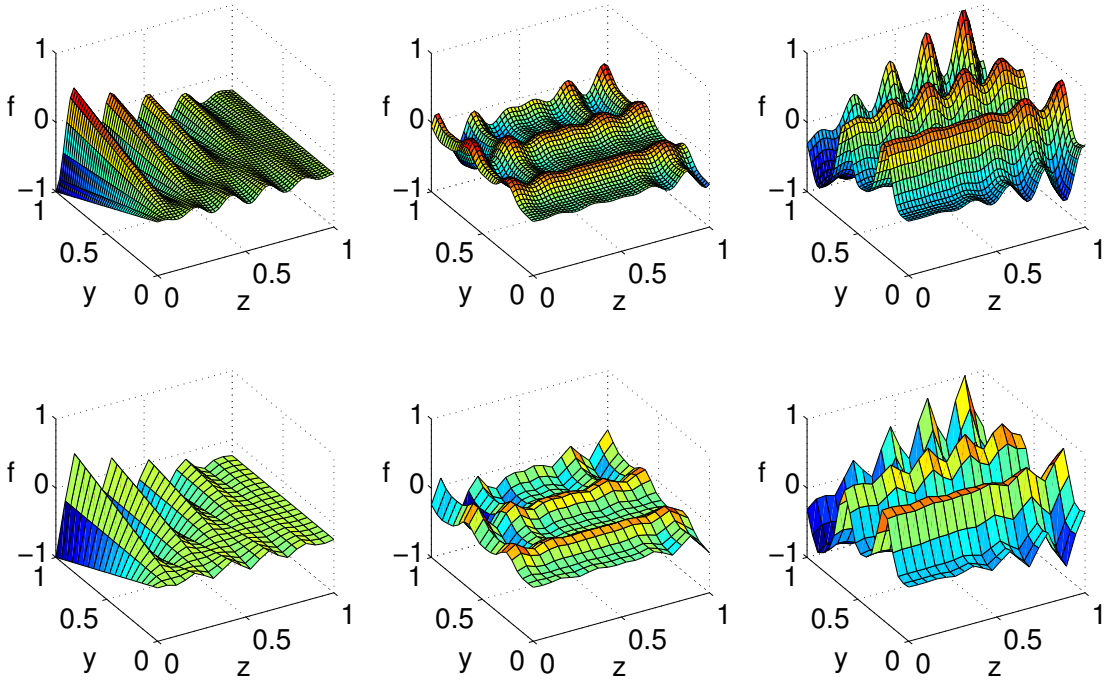

FIGURE 2.3: The dense (top) and coarse (bottom) toy model databases represented at constant x-value sections: $x = 0$ (left), $x = 1/2$ (middle) and $x = 1$ (right).

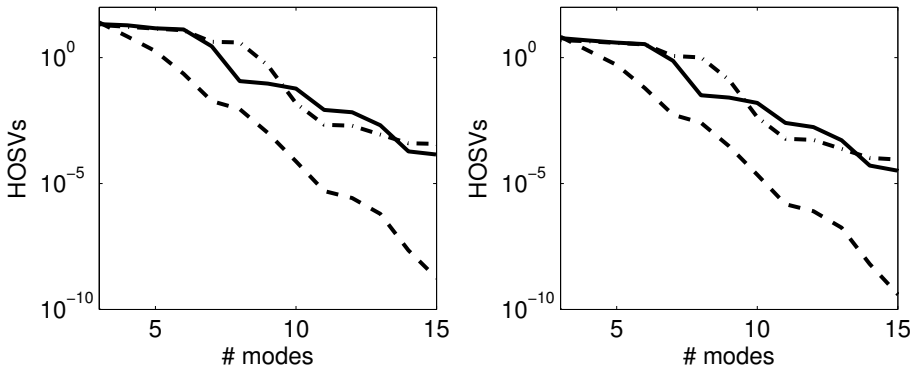

FIGURE 2.4: HOSVD singular values (HOSVs) associated with the indexes i (solid lines), j (dashed lines), and k (dot-dashed lines) when applying HOSVD to the dense (left) and coarse (right) toy model tensors.

dense and coarse toy model tensors are given in Figure 2.4. Note that the singular values distributions are quite similar in the dense and coarse meshes, which means that the redundancies along the three directions of the tensor are already well captured in the coarsest mesh. Also note that singular values in the index j decrease faster than those for the other two indexes. This is because the function f exhibits less structure in the y direction (see Figure 2.2).

Truncated HOSVD is illustrated in Figure 2.5, where both the RMS and maximum errors (denoted hereafter as RMSE and MaxE respectively),

$$\text{RMSE} = ||A - A^{\text{trunc}}||_{\text{RMS}}, \quad \text{MaxE} = \max_{ijk} |A_{ijk} - A^{\text{trunc}}_{ijk}|, \tag{2.14}$$

are plotted vs. the retained number of modes, $m + p + q$, for both the dense and coarse toy model databases. The error bound EB defined in Eq. (2.8) is also plotted for reference and seen to be fairly close to the RMSE. Note that

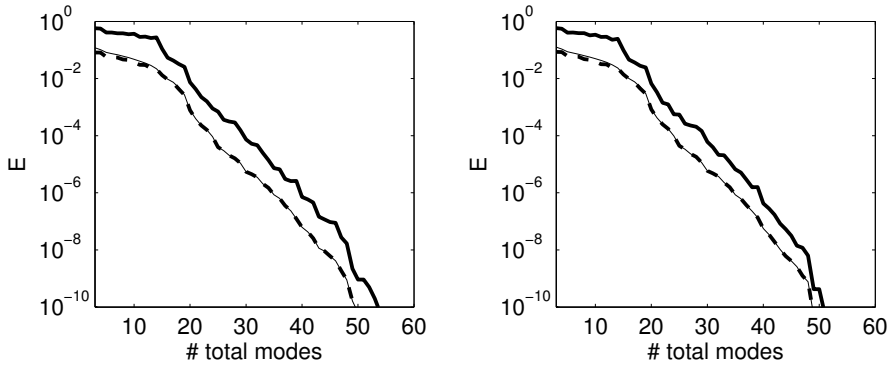

FIGURE 2.5: RMSE (dashed lines) and MaxE (thick solid lines) vs. the retained number of modes applying truncated HOSVD to the dense (left) and coarse (right) toy model databases. EB (defined in Eq. (2.8)) is also plotted for reference (thin solid lines).

both RMSE and MaxE decay slowly for $m + p + q \leq 14$, but as more modes are retained, both MaxE and RMSE decay spectrally. Also note that the method works equally well in the coarse and dense databases, which is a very good news for industrial applications, where the required meshes are usually fairly coarse. The method works fairly well since fixing a maximum acceptable error, e.g. of 10^{-4}, the required numbers of modes is 30 for both the coarse and dense databases.

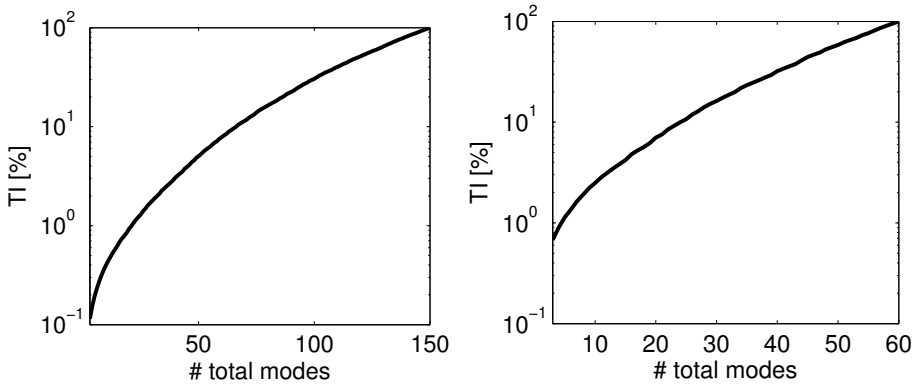

FIGURE 2.6: Truncated Information, calculated as in Eq. (2.10) vs. the retained number of modes applying truncated HOSVD to the dense (left) and coarse (right) toy model databases.

Maintaining 10^{-4} as the maximum acceptable error mentioned above, the truncated information when 30 modes are retained is $\sim 16\%$ and $\sim 2\%$ for the coarse and dense respectively. It demonstrates the efficiency of HOSVD decomposition in managing the most significant information with half or even less of the total modes. Note that the truncated information does not increase linearly with the total number of retained modes, as shown in Figure 2.6, but in a monotone convex way. Hence, TI growth at first third of retained modes is higher than last third. It means that HOSVD-based methods are efficiently suited for data reconstruction and treatment of databases, in general, dealing with the inherent patterns of the database. Furthermore, they do not need to know much information a priori about the nature of the database.

Chapter 3

Gappy reconstruction via iterated HOSVD

This chapter describes the theoretical definition of the techniques developed as the core of the present dissertation. Each method will be applied to the toy model databases, already described in §2.3, with the intention of illustrating and analysing their performance. Firstly, the ES-BR-VK method, already mentioned, will be recalled in order to easily understand its extension to high order.

The ES method (and the BR-VK extension) applies to 2D gappy databases A_{ij} and relies on POD. Here, it is described in terms of SVD, which facilitates its extension to higher dimensions. The idea in the ES method is to iterate the SVD reconstruction at gappy points, keeping the exact value of A_{ij} at the remaining points, which will be referred to as *non-gappy points*. More precisely, the method proceeds by iteratively applying truncated SVD, retaining a fixed number of modes, as follows:

- *First step*: An initial guess, A^0, is defined.
- *s-th step* (for $s \geq 1$): Truncated SVD is applied to the modified matrix $\hat{A}^s$ defined such that: (i)$\hat{A}^s = A^{s-1}$ at gappy points, where A^{s-1} corresponds to the truncated SVD matrix obtained at the (s-1)-th step, and (ii) $\hat{A}^s = A$ at non-gappy points.

As already mentioned in §1 and pointed out in [32], this iterative method depends on the initial guess and its performance may degrade if the number of retained modes is increased too much. This difficulty was solved in the BR-VK extension [33, 34] by using a second iteration on the number of retained modes. The resulting *doubly iterative method* is as follows:

- *First step:* The ES method is applied with just one mode, using as initial condition a guess of the gappy data.
- *m-th step:* The ES method is applied retaining $m > 1$ modes, taking as initial condition the converged result retaining $m - 1$ modes.

This method is fairly independent of the initial guess and behaves fairly monotonously [33, 34], at least in the absence of noise. Thus, it is this doubly iterative method that is extended to higher dimensions in the next section, where the treatment of noisy databases will also be addressed.

3.1 The gappy-HOSVD method

The extension of the doubly iterative ES-BR-VK method to higher dimensions can be made replacing standard SVD by HOSVD. For the sake of clarity, the method will be formulated in terms of a third order tensor but its extension to higher order tensors is straightforward. In addition, appropriate criteria are introduced for (i) the termination of the inner iteration in the ES method and (ii) the selection (among the three possibilities) of the HOSVD-mode that has to be added after each outer iteration. More precisely, beginning with an initial reconstruction (i.e., gappy elements are set to zero) of the gappy tensor A

$$A^0 = A \quad \text{and} \quad A^0_{ijk} = 0, (i,j,k) \in \{gappy\ points\}, \tag{3.1}$$

the *gappy-HOSVD method* is as follows:

- *Inner iteration:* Truncated HOSVD (for fixed numbers of retained modes along the three directions) is applied iteratively (for $s = 1, 2, \ldots$) to the modified tensor defined as $\hat{A}^s = A^{s-1}$ at gappy points and $\hat{A}^s = A$ at non-gappy points. The inner iteration is terminated if either

$$\text{(a) } EI_s < CF \cdot EB_s \tag{3.2}$$

or

$$\text{(b) } s > s_{\min}, \quad EI_s > EI_{s-1}, \text{ and } EI_{s-1} > EI_{s-2}, \tag{3.3}$$

where the *s-2 step* solution is considered the converged solution.

Here, the parameters CF and $s_{\min}$ are tunable, EB_s is the error bound defined in Eq. (2.8) for the truncated HOSVD at step s, and

$$EI_s = \sqrt{\frac{1}{|S^{gappy}|} \sum_{(i,j,k) \in S^{gappy}} |\hat{A}^s - \hat{A}^{s-1}|^2}, \tag{3.4}$$

where S^{gappy} is the set of indexes corresponding to the gappy points and $|S^{gappy}|$ is the total number of gappy points.

In other words, the inner iteration is terminated when either:

(a) The RMS difference (labeled as EI_s) between the reconstructions in the gappy part of the database in two consecutive iterations is a fraction of the EB. This is consistent with the fact that it would be useless

to achieve convergence beyond the precision of the HOSVD reconstruction. The convergence factor CF must be somewhat small and is tunable.

(b) The quantity EI_s reaches a minimum as s increases. Stopping the iteration at this minimum prevents divergence as s increases further, which would occur in some cases. Hence, the converged solution of the inner iteration when this condition (3.3) is fulfilled, corresponds to the *s-2 step*. The minimum number of iterations in (3.3) avoids a possible irregular behaviour of the iterative process in the initial iterations. Although $s_{\min}$ could be tunable, it is recommended to be set as the number of dimensions plus one. The performance of the method is fairly insensitive to changes of $s_{\min}$ considering it small.

- *Outer iteration:* The inner iteration is repeated, beginning with $m = p = q = 1$ and adding one HOSVD-mode at a time, taking as initial guess the converged solution in the previous iteration. Since there are three types of HOSVD-modes (one per index), three possible modes can be selected to be added for the next iteration. A convenient selection is that mode promoting the largest decrease in EB (see Eq. (2.8)). Thus, if the numbers of modes in the current inner iteration in the directions i, j, and k are m, p, and q, respectively, then the maximum among the HOSVD singular values α_{m+1}, β_{p+1}, and γ_{q+1} is calculated and the associated HOSVD-mode is selected for the next inner iteration.

Now, since the data provided (at non-gappy positions) are considered as exact, the reconstruction is performed using the provided data at non-gappy points and the HOSVD-reconstruction at gappy points.

3.1.1 The gappy-HOSVD algorithm

Thus far, the proposed method is theoretically explained. But some practical aspects should be considered in order to implement it in a numerical algorithm.

The convergence conditions (3.2)-(3.3) have been tested on the toy model databases. The conclusion is that, with these conditions, the inner iteration behaves as expected. Namely, in the inner iterations associated with the first few outer iterations, EB is only moderately small and the convergence conditions behave somewhat irregularly, namely some inner iterations are terminated by the condition (3.2) and some others by (3.3). But, as progressing in the outer iterations, the error bound (EB) appearing in (3.2) becomes appropriately small, the HOSVD approximation behaves more regularly, and it is condition (3.2) that terminates the inner iteration.

When the number of retained modes becomes higher, EI and EB adopt values near the machine precision generating abnormal behaviour in order to fulfil the exit criteria and stopping the inner loop. Thus, the number of inner iterations increases considerably without improvement of the reconstruction. Enclosing

those variables within limits avoids unnecessary iterations of the inner reconstruction when the HOSVD approximation is sufficiently accurate for a fixed number of total modes. Specifically, EB should be delimited by a maximum and a minimum, set as $EB_{max} = 1$ and $EB_{min} = 10^{-8}$ for all tested cases in this Thesis. For EI, a maximum value is set, i.e. 0.1. These bounds serve to accelerate the algorithm without affecting the global performance of the method. Both toy model databases have been tested with and without those limits, reproducing a wide range of different cases: (i) with randomly distributed gappyness, 50 samples per three different percentages (20%, 40% and 60%) and (ii) with interior concentrated black-box like gappyness, 5 different samples for three different percentages ($\sim$30%, $\sim$40% and $\sim$50%). This analysis tries to evaluate the behaviour of the bounds independently of the position and the type of the gappyness. Differences in RMSE and MaxE between gappy-HOSVD reconstructions with and without bounds in the inner loop parameters when $m + p + q > 40$ are less than 10^{-8} but the computing time increased more than 50% in all tested cases.

Once the bounded inner loop has been demonstrated to be more efficient, the tunable convergence factor CF has been analysed in the range $0.001 \leq CF \leq 0.1$, in order to determine its influence on the method performance. The gappy-HOSVD method has been iteratively applied to a set of several samples with both randomly distributed (50 samples per each studied percentage) and concentrated black-box-like (5 samples per each studied percentage) gappyness. An exhaustive comparison is shown in Tables 3.1 and 3.2 for the randomly distributed gappyness study and in Tables 3.3 and 3.4 for the black-box gappyness study. The lowest percentage of gappyness has been omitted with the intention of showing the worst scenarios (larger percentages). Concentrating on the former type of gappyness, errors are fairly insensitive to this parameter in comparison with the required CPU time, which values increase considerably when CF is very small. It should be noticed that the variance of MaxE values is almost zero, demonstrating the exceptional repetitiveness of the method. On the other hand, the concentrated black-box like gappyness study reveals slightly bigger differences in MaxE when the CF is smaller and also it is reflected in the required CPU time, which in some cases is of the order of hours.

Summarizing, a compromise decision should be taken between precision an computing time in order to assign a value to CF. In this Thesis $CF = 0.1$ has been selected as a good choice for all subsequent applications since the method gives results in a few seconds with acceptable errors level.

Furthermore, the number of steps in the inner iteration can be limited by $s_{\max}$. This parameter is also tunable and directly related with CF, since in practice, the stop condition (3.2) is fulfilled in a higher percentage of cases than (3.3). This condition allows to accelerate the convergence speed of the method although in some cases the optimum reconstruction is not achieved. A practical study about the computing time in relation to $s_{\max}$ when $CF = 0.1$ has been done with the same samples set as before. Limiting the maximum number of iterations to 50 yields complete convergence (according to Eqs. (3.2)-(3.3)) in both dense

	40%			60%		
CF	**t [sec.]**	**# iter**	**MaxE**	**t [sec.]**	**# iter**	**MaxE**
0.100	3	7	$8.25 \cdot 10^{-3}$ [0.0%]	4	9	$1.12 \cdot 10^{-2}$ [0.1%]
0.050	7	11	$8.02 \cdot 10^{-3}$ [0.0%]	12	20	$9.98 \cdot 10^{-3}$ [0.0%]
0.020	11	19	$7.91 \cdot 10^{-3}$ [0.0%]	24	48	$9.42 \cdot 10^{-3}$ [0.0%]
0.001	59	106	$7.88 \cdot 10^{-3}$ [0.0%]	340	641	$9.16 \cdot 10^{-3}$ [0.0%]

TABLE 3.1: Statistical analysis of the CF parameter influence of the gappy-HOSVD method applied to the coarse toy model database with different percentages of randomly distributed gappyness. The required CPU time (t) and the number of inner iterations (iter) represent mean values of 50 samples. MaxE represents the mean and the variance (as percetage of the mean) of the maximum error obtained for all 50 samples in the gappy region using $m + p + q = 20$ modes.

	40%			60%		
CF	**t [sec.]**	**# iter**	**MaxE**	**t [sec.]**	**# iter**	**MaxE**
0.100	27	4	$7.60 \cdot 10^{-3}$ [0.0%]	32	5	$8.17 \cdot 10^{-3}$ [0.0%]
0.050	27	5	$7.58 \cdot 10^{-3}$ [0.0%]	39	7	$7.71 \cdot 10^{-3}$ [0.0%]
0.020	41	6	$7.46 \cdot 10^{-3}$ [0.0%]	61	10	$7.90 \cdot 10^{-3}$ [0.0%]
0.001	85	14	$7.61 \cdot 10^{-3}$ [0.0%]	137	26	$7.78 \cdot 10^{-3}$ [0.0%]

TABLE 3.2: Counterpart of Table 3.1 for the dense toy model database.

	38%			49%		
CF	**t [sec.]**	**# iter**	**MaxE**	**t [sec.]**	**# iter**	**MaxE**
0.100	5	11	$1.28 \cdot 10^{-2}$ [0.2%]	5	10	$2.17 \cdot 10^{-2}$ [0.0%]
0.050	11	25	$8.94 \cdot 10^{-3}$ [0.1%]	14	30	$1.87 \cdot 10^{-2}$ [0.1%]
0.020	35	78	$5.96 \cdot 10^{-3}$ [0.1%]	38	84	$1.42 \cdot 10^{-2}$ [0.3%]
0.001	537	1192	$4.41 \cdot 10^{-3}$ [0.0%]	760	1662	$7.58 \cdot 10^{-3}$ [0.2%]

TABLE 3.3: Statistical analysis of the CF parameter influence of the gappy-HOSVD method applied to the coarse toy model database with different percentages of concentrated black-box gappyness. The required CPU time (t) and the number of inner iterations (iter) represent mean values of 5 samples. MaxE represents the mean and the variance (as percetage of the mean) of the maximum error obtained for all 5 samples in the gappy region using $m + p + q = 20$ modes.

	39%			50%		
CF	**t [sec.]**	**# iter**	**MaxE**	**t [sec.]**	**# iter**	**MaxE**
0.100	44	10	$1.45 \cdot 10^{-2}$ [0.1%]	44	10	$1.96 \cdot 10^{-2}$ [0.0%]
0.050	123	27	$1.04 \cdot 10^{-2}$ [0.2%]	130	26	$1.77 \cdot 10^{-2}$ [0.0%]
0.020	498	84	$6.82 \cdot 10^{-3}$ [0.1%]	568	101	$1.33 \cdot 10^{-2}$ [0.1%]
0.001	6036	1098	$5.09 \cdot 10^{-3}$ [0.0%]	9146	1639	$7.07 \cdot 10^{-3}$ [0.0%]

TABLE 3.4: Counterpart of Table 3.3 for the dense toy model database.

and coarse databases using an average number of iterations less than 15, which required $\sim$ 40 CPU seconds, for all studied gappyness types and percentages. However, extending this study to samples where the black-box gappyness is next to the boundary of the domain and therefore the reconstruction involves extrapolation, some differences in convergence have been found. In this case (both dense and coarse databases) convergence was achieved instead in about 70% of the outer iteration steps, and required 2 CPU minutes.

Increasing the number of allowed iterations in this particular case study to 1,000, approximately reduces the maximum errors in the gappy region from MaxE=$4.6 \cdot 10^{-2}$ to MaxE=$3.2 \cdot 10^{-2}$ (with 23 retained modes) and from MaxE=$6.0 \cdot 10^{-2}$ to MaxE=$5.9 \cdot 10^{-2}$ (with 22 retained modes) for the dense and coarse databases, respectively. These differences, almost zero in the coarser database, confirms that convergence was not achieved with 50 iterations, but the required CPU time has increased almost 30%. This means that convergence with an unlimited number of iterations in the inner loop is fairly slow in highly demanding databases, as it happened with the original ES method.

3.1.2 The gappy-HOSVD method application

The gappy-HOSVD method is now illustrated for the two toy model databases defined in §2.3. The tunable parameters of the method are taken here as $CF = 0.1$ and $s_{\max} = 50$. The performance of the method will be tested in terms of RMSE and MaxE, defined in (2.14) for the whole database; the counterparts of RMSE and MaxE in the gappy and non-gappy parts of the database are defined in a similar way and will also be considered.

For *randomly distributed gappyness*, Figure 3.1 shows that the reconstruction is very good for the considered gappyness levels: the RMSE and MaxE curves for the reconstructed gappy database (applying the gappy-HOSVD method) and the reconstructed exact database (applying truncated HOSVD), retaining the same numbers of modes along the three directions, overlap up to a certain number of retained modes; thus, the reconstruction is somewhat optimal in this case. As expected, this number of retained modes is smaller for the coarse database than for the dense database and decreases as the level of gappyness increases. The method is very efficient in the dense database, where the

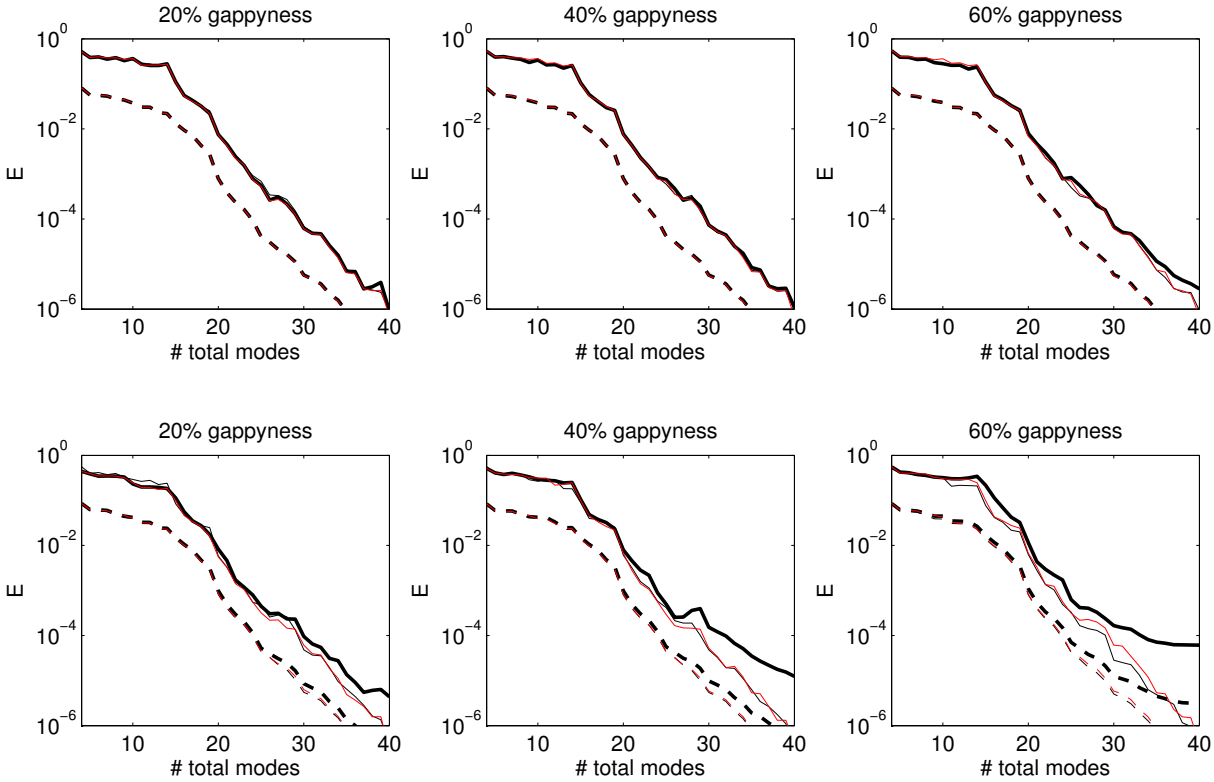

FIGURE 3.1: Application of the gappy-HOSVD method to the dense (top) and coarse (bottom) toy model databases, plotting RMSE (dashed lines) and MaxE (solid lines), as defined in Eq. (2.14), of the reconstructed solution in the gappy (thick-black) and the non-gappy (thin-black) regions. For reference, the result of applying truncated HOSVD (retaining the same numbers of modes) to the clean databases are also plotted with red lines.

errors steadily decrease as the number of retained modes increases, even for 60% gappyness. Obviously, the reconstructions are not as good in the coarse database, but still the 60% missing data are reconstructed with a MaxE $\sim 10^{-2}$ and $\sim 10^{-4}$ retaining 20 and 30 modes, respectively, which is remarkable taking into account how coarse the coarse database is. Figure 3.2 shows an example of the reconstruction when gappyness is 60% for both databases in a section. Clearly, errors cannot be appreciated by the eye when those reconstructions are compared with their counterparts for the clean databases, see Figure 2.2.

Concentrated gappyness is, of course, more demanding, but is also efficiently reconstructed by the method. For illustration, we consider the following gappy regions in the coarse toy model database

$$\text{(a): } 0.15 < x, y < 0.85;\ \forall(z) \quad \text{and} \quad \text{(b): } (x-0.5)^2 + (y-0.5)^2 > 1.2;\ \forall(z), \tag{3.5}$$

namely, (a) the inside of a square cylinder and (b) the outside of a circular cylinder, respectively. Likewise, two gappy regions have been considered in

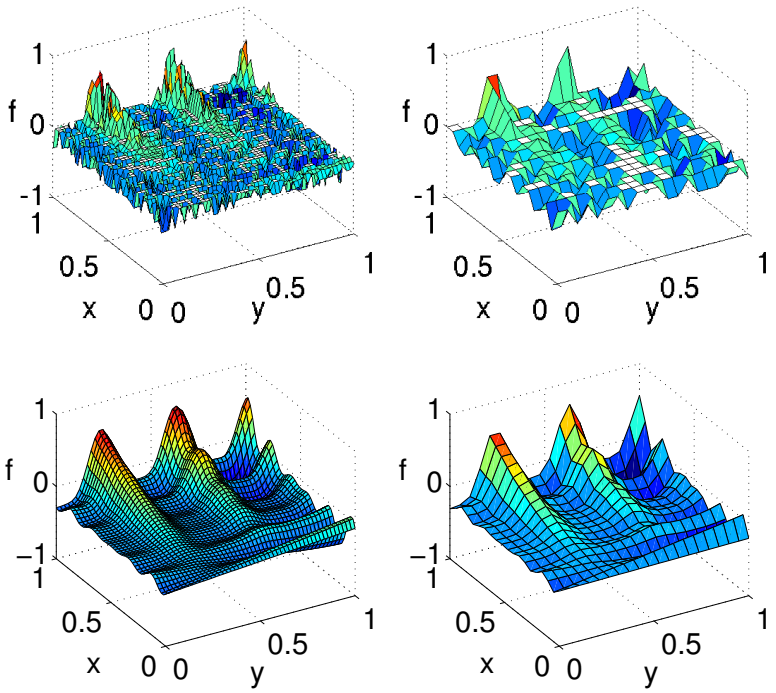

FIGURE 3.2: $z = 0.5$-cross-section of the dense (left) and coarse (right) toy model databases with 60% gappyness (top) and their reconstruction (bottom) using $m + p + q = 20$ modes.

the dense toy model database

$$\text{(a): } 0.14 < x, y < 0.86;\ \forall(z) \quad \text{and} \quad \text{(b): } 0.14 < x < 0.86;\ 0 \leq y < 0.72;\ \forall(z), \tag{3.6}$$

where both are the same square cylinder (a) centred in the domain and (b) shifted towards the boundary of the domain. Note that the gappy regions are black-box, namely they include all values of z for each considered (x, y). To illustrate the level of gappyness, the sections $z = 0.5$ for the gappy databases are plotted in Figures 3.3 and 3.4 , where the gappy points are easily identified because the value A=0 is assigned to these points. The counterparts of the

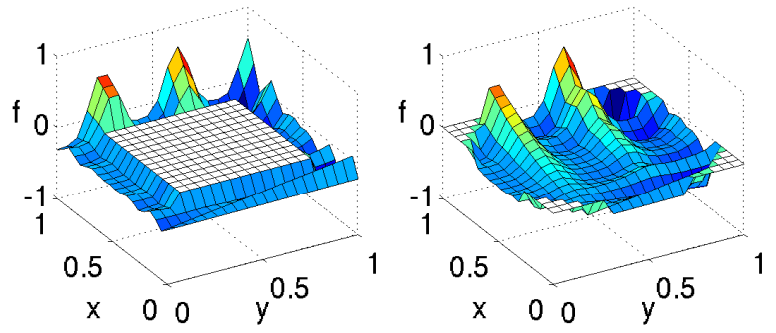

FIGURE 3.3: $z = 0.5$-cross-section of the coarse toy model database with gappy regions defined in Eq. (3.5).

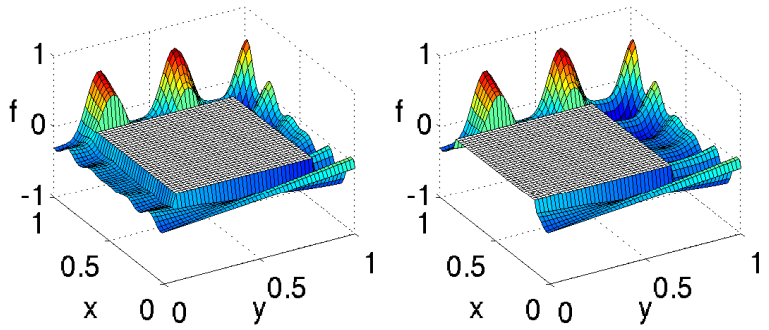

FIGURE 3.4: $z = 0.5$-cross-section of the dense toy model database with gappy regions defined in Eq. (3.6).

plots in Figure 3.1 for the coarse database with black-box gappyness are given in Figure 3.5; for the dense database with square gappy regions as defined in Eq. (3.6) are shown in Figure 3.6.

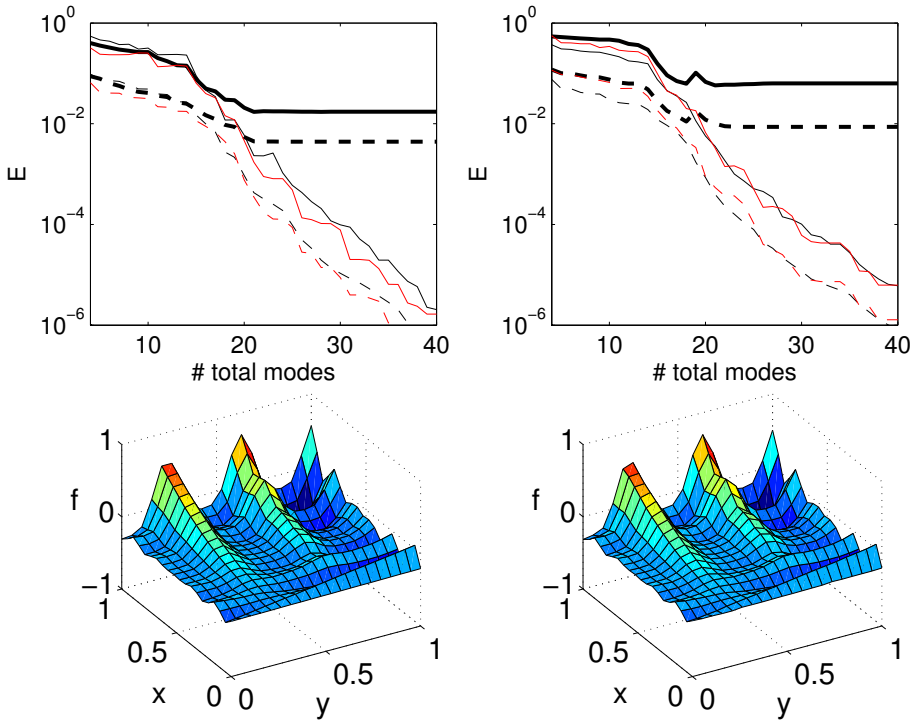

FIGURE 3.5: Counterparts of Figure 3.1 for the coarse gappy database, with gappy regions defined in Eq. (3.5a) (left) and Eq. (3.5b) (right). $z = 0.5$-cross-section of the coarse toy model database reconstructions using $m + p + q = 21$ modes (left) and $m + p + q = 18$ modes (right).

Figure 3.5-left-top is qualitatively similar to its counterparts in Figure 3.1. In fact, it is remarkable that retaining 21 modes, the errors in the reconstructed gappy region are RMSE$= 4 \cdot 10^{-3}$ and MaxE$= 1.7 \cdot 10^{-2}$, in spite of the fact that the database is quite coarse and that gappy reconstruction requires a very

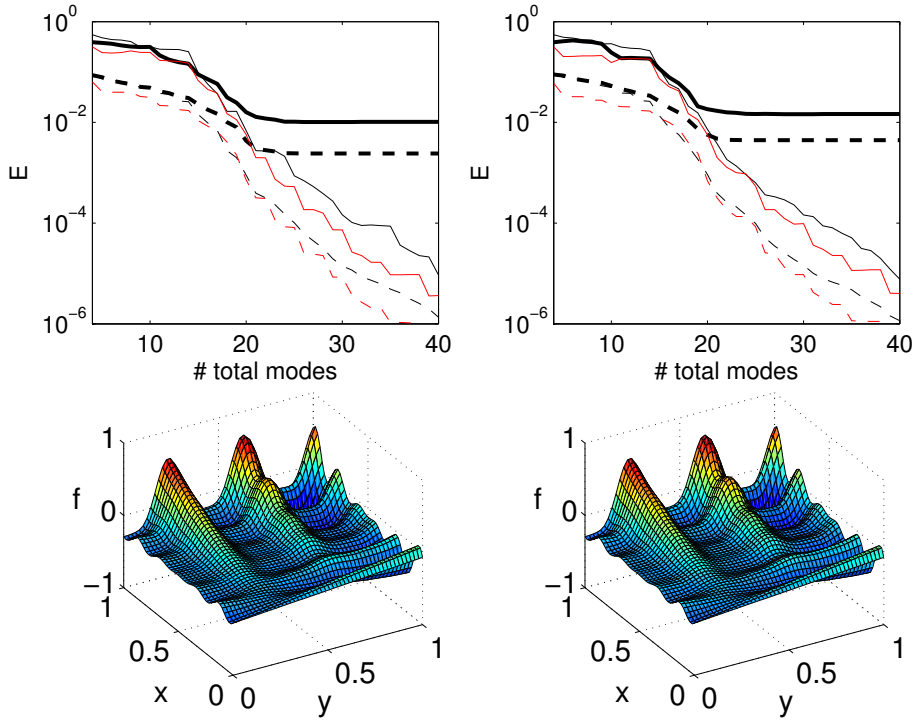

FIGURE 3.6: Counterparts of Figure 3.1 for the dense gappy database, with gappy regions defined in Eq. (3.6a) (left) and Eq. (3.6b) (right). $z = 0.5$-cross-section of the dense toy model database reconstructions using in both cases $m+p+q = 24$ modes.

dramatic global interpolation: only three database shells are available at the lateral sides to reconstruct the gappy region, see Figure 3.3.

Same error trends can be observed for the dense database in Figure 3.6. The gappy region (3.6a) is slightly bigger than its counterpart for the coarse database (supposes 52% of total points in comparison with 49% that belongs to the coarser) yet RMSE reduces by half: MaxE=$1.0 \cdot 10^{-2}$ and RMSE=$2.4 \cdot 10^{-3}$ when retaining 24 modes. The reconstruction plot in Figure 3.6-left-bottom is clearly similar to its counterpart of the 'clean' database, in Figure 2.2. Same parallelepiped gap is considered in the Figure 3.6-right but shifted. Now reconstruction involves *extrapolation along the y-axis* (Figure 3.4-right). In spite of that, the errors in the final reconstruction with $m + p + q = 24$ modes are, MaxE=$1.5 \cdot 10^{-2}$ and RMSE=$4.4 \cdot 10^{-3}$, similar to those obtained in the coarse database when the gappy region involves strict interpolation.

Figure 3.5-right shows a worse reconstruction in the gappy region than in the left plot, which is due to the fact that gappy reconstruction involves now an *extrapolation* to the outside of the circular cylinder, while the gappy reconstruction in the left plot involved interpolation. The reconstructions, shown in Figure 3.5-bottom, are hardly distinguishable to eye between them, and also in comparison with their counterpart of the clean database in Figure 2.2. But it is interesting that the HOSVD reconstruction in the non-gappy region is still quite

good (as is in the left plot), which suggests a possible application of this method to construct HOSVD representations in non-rectangular domains (such as the circular cylinder in the present case). This application will be further pursued in §4.2, where an aerodynamic database defined in a non-rectangular domain will be treated using this method.

Summarizing the above, the gappy-HOSVD method gives quite good reconstructions in the toy model database, both when gappyness is randomly distributed and when it is concentrated (including black-box gappyness), especially when gappy reconstruction involves interpolation. As expected, randomly distributed gappyness is less demanding. The case in which gappy points are located black-box-like is very interesting since it is a case that could not be treated using standard POD-based gappy reconstruction methods, as detailed in [6] where Kriging interpolation is proposed instead.

3.2 The gappy-noisy-HOSVD method

As mentioned, SVD-like decompositions enable to deal efficiently with situations where near rank deficiency prevails [23] as in noisy databases. As in the ES-BR-VK method [34], the presence of noise contaminates the quality of the available data and thus worsens the performance of the gappy-HOSVD method, especially for large gappyness. Because of noise, the error of the reconstruction does not decrease monotonously as the number of retained modes increases. Instead, the error first decreases and then increases after a threshold number of modes. First decreasing stage is expected because the HOSVD approximation is improved as more modes are added (and more and more contaminated positions are corrected). The problem is that if too many modes are retained, the highest order modes account for redundancies associated with the errors, not for the redundancies of the clean part of the database. A key observation is that when too many modes are added, then the redundancies associated with the errors contaminate all modes. Thus, a good criterion results from observing the difference between the modes (in all the directions of the database) in two consecutive outer iterations (associated with increasing the number of modes) in the gappy-HOSVD method. Such threshold, where minimum reconstruction errors occur, is estimated as that giving the minimum of the following *mode selection estimate*

$$\text{MSE} = \frac{\sqrt{\sum_{i'=1}^{m} \alpha_{i'}^2 ||U_{new}^{i'} - U_{old}^{i'}||^2}}{\sqrt{\sum_{i'=1}^{m} \alpha_{i'}^2}} + \frac{\sqrt{\sum_{j'=1}^{p} \beta_{j'}^2 ||V_{new}^{j'} - V_{old}^{j'}||^2}}{\sqrt{\sum_{j'=1}^{p} \beta_{j'}^2}} + \frac{\sqrt{\sum_{k'=1}^{q} \gamma_{k'}^2 ||W_{new}^{k'} - W_{old}^{k'}||^2}}{\sqrt{\sum_{k'=1}^{q} \gamma_{k'}^2}} \tag{3.7}$$

in the interval

$$N_{\min} \leq m + p + q \leq N_{\max}, \tag{3.8}$$

where N_{min} and N_{max} are tunable. Here, (using the same notation as in §2.2), $\alpha_{i'}$, $\beta_{j'}$ and $\gamma_{k'}$ are the HOSVD singular values associated with the fibers along the directions i, j and k, respectively, and $||\cdot||$ denotes the usual vector (Euclidean) norm. The modes numbers m, p and q are those retained in the former (outer) iteration, and the subscripts old and new denote the HOSVD-modes in the former and current (outer) iterations, respectively.

As anticipated, the idea behind the MSE defined in (3.7) is that if noise is small and exhibits redundancies that are very different from those in the clean database, which is reasonable, then the most energetic modes are associated with true redundancies of the clean database, while noise is associated with the less energetic modes. This is due to the fact that SVD-based decompositions (including higher order decompositions), as the proposed method represents, are able to indicate how near a given tensor is to a tensor of lower rank [23], which is nearer to the exact database. Thus, if only the most energetic 'clean' modes are retained, then the MSE and the error of the gappy-HOSVD reconstruction will both decrease as the number of retained modes increases. On the other hand, if too many modes are retained, the redundancies associated with the noise will come into play (in the reconstruction at non-gappy points that is performed at the inner iteration) and contaminate all retained modes, promoting an increase in both the MSE and the error of the gappy-HOSVD reconstruction. The crossover between these two situations will occur at the minimum of MSE, which is expected to give a good estimate of the 'optimal' numbers of retained modes. Restricting from below the interval (3.8) (through N_{min}) avoids the irregular behaviour when the number of retained modes is too small; the restriction from above (through N_{max}) avoids a possible spurious decrease of the MSE when a sufficient number of modes are retained as to also capture the redundancies associated with noise.

Clearly, N_{min} and N_{max} selection becomes an important issue when the gappy-noisy-HOSVD method is applied to a new database. A priori information about the database is not available in order to determine these values. However, the study cases presented throughout this Thesis, representing different nature of information, demonstrate the robustness of the method when same fixed values are used for all cases. N_{min} should be small. It is recommended to be setted as the number of dimensions plus two. This criterion keeps the method from irregular behaviour independently of the nature or noise size of the database. Independently of the database size, N_{max} is setted as 30 for all studied cases, except for the experimental database application where the total number of modes is smaller. In this single case, N_{max}=15 will be used instead.

Now, the reconstruction of the database retaining the 'optimal' number of modes can be made in two ways:

A. By using the given data at non-gappy points and the HOSVD reconstruction at gappy points, as in the plain gappy-HOSVD method. This reconstruction is consistent with (and will be compared below with) the contaminated database.

B. By using the HOSVD reconstruction at all points. This reconstruction is consistent with (and will be compared with) the clean database.

The RMSE and MaxE errors calculated using these two methods will be denoted as

$$\mathrm{RMSE}^{\mathrm{cont}}, \mathrm{MaxE}^{\mathrm{cont}} = \text{calculated according to item A}, \tag{3.9}$$

$$\mathrm{RMSE}^{\mathrm{clean}}, \mathrm{MaxE}^{\mathrm{clean}} = \text{calculated according to item B}. \tag{3.10}$$

The core method described above will be referred to as the *gappy-noisy-HOSVD method*. Both options, A and B, become variants of this method which application depends on the needs.

3.2.1 The gappy-noisy-HOSVD method application

Evaluating both strategies (A and B), mentioned above, is only possible when both the clean and the contaminated databases are known, as occurs in the toy model databases, where the exact values are known and the noise is added. It is going to be used to elucidate the performance of the method in terms of gappy reconstruction and noise filtering.

In order to illustrate its performance, both dense and coarse toy model databases are going to be contaminated with random, uniformly distributed, positive noise of size 0.05 at all positions of the databases.

In the following, *noise size* refers to as the maximum magnitude of noise present in the database. We remind that the maximum value of the clean toy model database is 1. Hence, adding a positive noise of size 0.05 means that the noise adopts a value lower than 5% of the maximum value of the clean database.

The tunable parameters appearing in (3.8) are taken as $N_{\mathrm{min}} = 5$ and $N_{\mathrm{max}} = 30$.

The counterpart of Figure 3.1 is given in Figure 3.7, where the RMSE and MaxE are calculated according to Eqs. (3.9)-(3.10) and the mode selection estimate, MSE, is also plotted. Note that errors resulting from comparing with the contaminated and clean databases exhibit similar trends.

As anticipated, errors in the gappy part of the databases first decrease and then increase, reaching a minimum that is reasonably well predicted by the MSE. For instance, for 60% gappyness in the coarsest database, the optimal numbers of modes are $(m+p+q)_{optim} = 8+5+8 = 21$, where RMSE$= 8.1 \cdot 10^{-3}$ and MaxE$= 2.9 \cdot 10^{-2}$ in the gappy region. The estimate MSE reaches its minimum nearby, at $m+p+q = 8+7+9 = 24$, where RMSE$= 8.4 \cdot 10^{-3}$ and MaxE$= 3.3 \cdot 10^{-2}$, which are fairly close to the optimal values. In the case of the dense database, the performance of the method is better, as expected. Errors are slightly lower and MSE predicts fairly well the optimal number of modes in all three percentages of studied gappyness. The reconstructions according to the strategy B (3.10) for both coarse and dense databases are shown in Figure 3.8-bottom. Figure 3.8-top

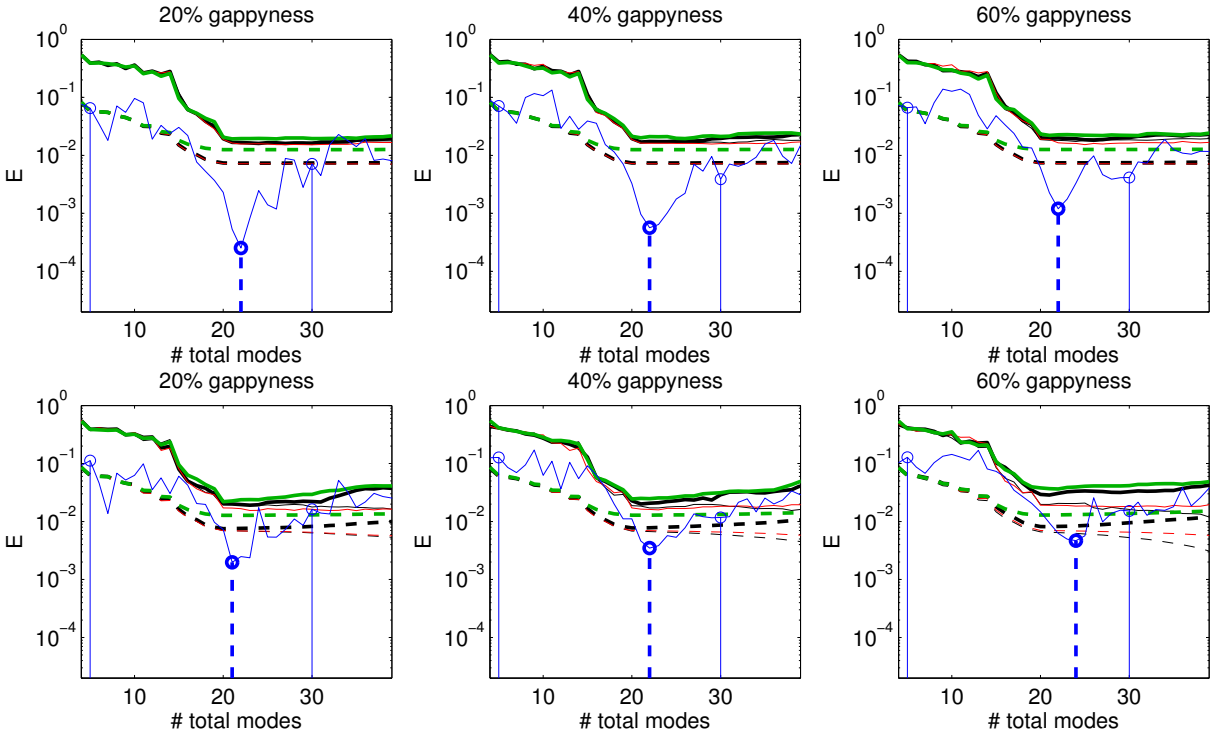

FIGURE 3.7: Counterpart of Figure 3.1 applying the gappy-noisy-HOSVD method to the dense (top) and coarse (bottom) toy model databases contaminated with random, uniformly distributed, positive noise of size 0.05; the mode selection estimate defined in Eq. (3.7) is plotted with blue line. The errors $(\mathrm{RMSE}^{\mathrm{cont}}, \mathrm{MaxE}^{\mathrm{cont}})$ and $(\mathrm{RMSE}^{\mathrm{clean}}, \mathrm{MaxE}^{\mathrm{clean}})$ (defined in Eqs. (3.9)-(3.10)) for the gappy region are plotted with thick black and green lines, respectively.

which are cross sections of the contaminated databases (coarse and dense toy model ones) with 60% gappyness, clearly exhibits the reconstruction challenge that the combination of noise and high level of gappyness represents.

It is very important to note that the $\mathrm{MaxE}^{\mathrm{clean}}$ is always smaller than the added noise since $\mathrm{MaxE}^{\mathrm{clean}}=3.9 \cdot 10^{-2}$ in the worst scenario, 60% gappyness for the coarsest database. It means that the method filters a part of the noise out in spite of the fact that the added noise is always positive, namely it does not exhibit a zero mean.

The method works also well with concentrated gappyness. In particular, for the counterparts of Figure 3.5 with a random, positive noise of size 0.05, the optimal numbers of modes are well captured and the reconstruction is fairly good, as Figure 3.9 shows. $\mathrm{MaxE}=3.5 \cdot 10^{-2}$ and $\mathrm{RMSE}=1.0 \cdot 10^{-2}$ are obtained when the gappy region defined in Eq. (3.5a) is present, and $\mathrm{MaxE}=6.1 \cdot 10^{-2}$ and $\mathrm{RMSE}=1.1 \cdot 10^{-2}$ in the case of Eq. (3.5b). Note that also in these cases, the noise is inherently filtered out by the method since $\mathrm{MaxE}^{\mathrm{clean}}$ is lower than the noise size, except for the case that requires an extrapolation. The counterpart of Figure 3.6 with the same noise level (random positive noise of size 0.05),

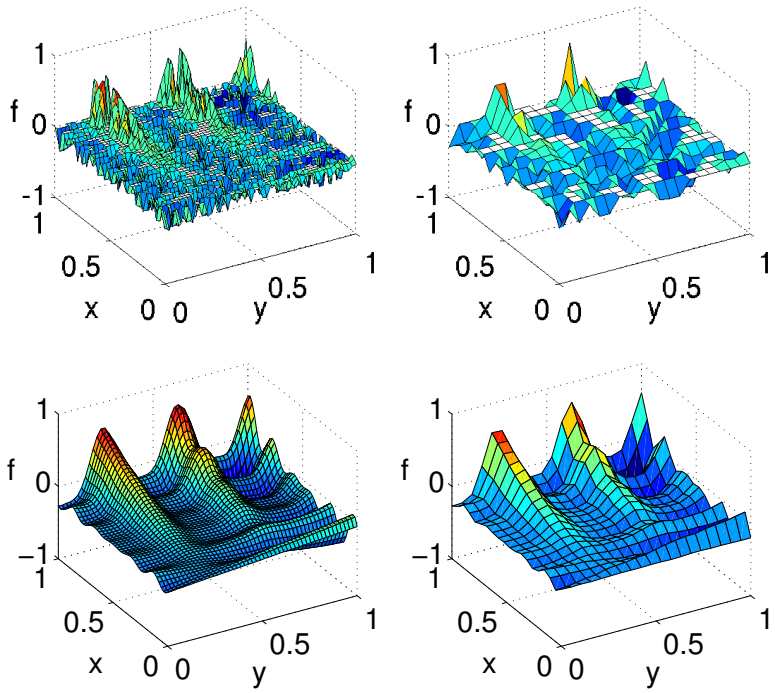

FIGURE 3.8: Counterpart of Figure 3.2 when positive noise of size 0.05 is present. The reconstruction uses $m + p + q = 22$ and $m + p + q = 24$ modes respectively from left to right, according to MSE estimation.

exhibits also good performance in reconstructing missing data considering how big the gappy region is. The error and reconstruction plots are similar to those in Figure 3.6.

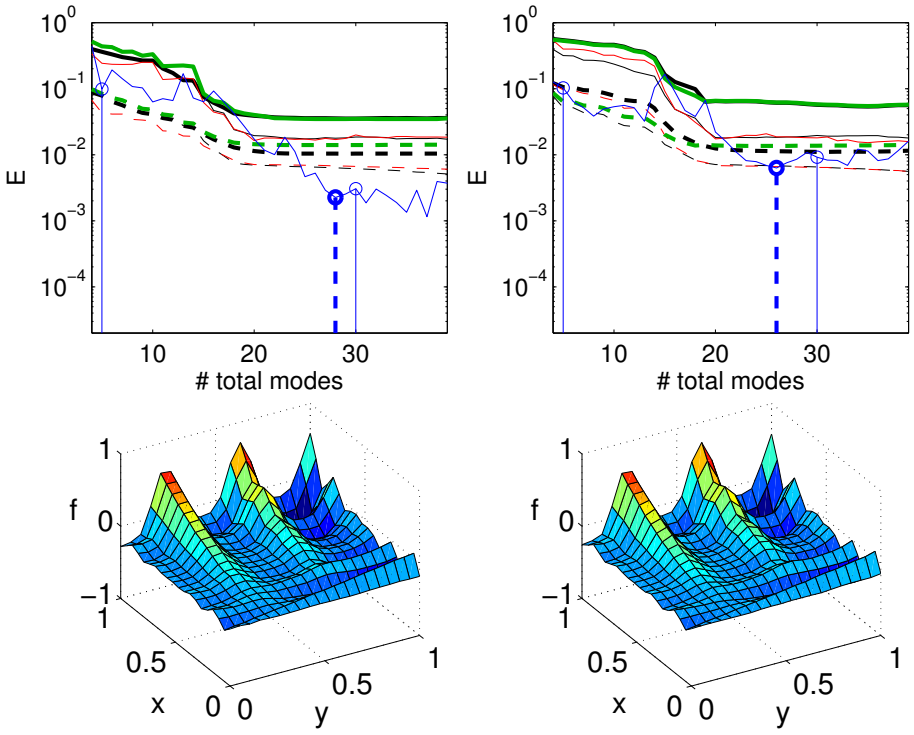

FIGURE 3.9: Counterpart of Figure 3.5 for the coarse gappy database when positive noise of size 0.05 is present, with gappy regions defined in Eq. (3.5a) (left) and Eq. (3.5b) (right). The reconstructions are plotted using $m+p+q=28$ and $m+p+q=26$ modes respectively from left to right, according to MSE estimation.

Chapter 4

Application to a theoretical aerodynamic database

The methods developed in the previous chapter are now applied to theoretical aerodynamic databases that exhibit most of the difficulties expected from a real case but with the advantage that, being theoretical, the incidence of errors is controllable. The aim is to study the behaviour of these methods with different mesh definitions as same as evaluating the incidence of diverse grid aspects in the reconstruction quality. These databases can be considered as 'exact' and experimental noise is artificially added. Therefore, they should be generated by a theoretical function implemented for evaluating the studied physics exactly at the desired mesh points. Once the theoretical databases are calculated in §4.1, the mathematical tools can be applied and the results are presented in §4.2. This section includes direct applications of the methods on the complete database and strategies to be applied on portions of these databases.

4.1 Three-dimensional database generation

This section intends to describe the generation of the mentioned theoretical databases based on the calculated pressure distribution on a wing. This aerodynamic problem is of great interest to the industry and is usually determined after wind tunnel testing. Starting from a two dimensional analysis in Appendix A, the distribution of pressure over an airfoil which easily allows to study its aerodynamic performance, it can be generalized over the span of the wing in order to obtain a pseudo-three-dimensional pressure distribution. This information can be stored in a matrix where the local chordwise and spanwise coordinates of the wing surface geometry are the independent variables and varying the angle of attack one obtains the third dimension of the database.

The bidimensional analysis of pressure coefficient distribution on an airfoil has been widely addressed in literature. A well-known semi-empirical method by

Theodorsen (detailed in Appendix A) has been carefully implemented for calculating the inviscid, potential flow past airfoil. This bidimensional tool is applied spanwise of a 3D wing model in order to obtain the complete C_p distribution over the wing.

The bidimensional analysis is able to exactly evaluate any point of the wing mesh because it is previously included in the airfoil discretization. This correction avoids the use of interpolation but forces to call this tool iteratively. Furthermore, the airfoil discretization used at each iteration is similarly defined in terms of spacing and number of points in order to avoid malfunctioning of the inverse mapping and minimize the numerical dicretization errors of this tool.

Strong 3D effects near the wing tip are accounted for near the 2% of the tip edge using a logarithmic correction forcing zero C_p values at the tip.

This theoretical function allows to create aerodynamic toy model tensors whose elements are evaluated at any desired point of the parameter space.

A sample wing is chosen with an elliptical planform (see Figure 4.1) whose

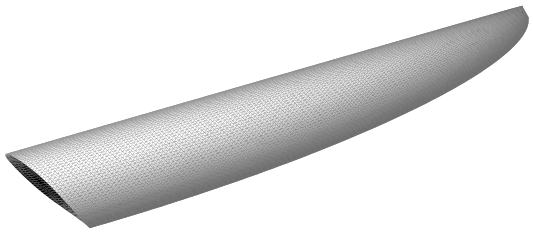

FIGURE 4.1: A three-dimensional view of a half of the elliptic planform wing.

airfoil is a NACA 64-210; the aspect ratio is 9. Note that the wing planform exhibits neither swept nor twist.

The three dimensional database is defined by three spatial variables: s-coordinate (defined as described in section §A), y-coordinate (which increases from the root section, $y = 0$, to the wing tip, $y = y_{wt}$) and the angle of attack. First two variables represent the geometrical coordinates of the grid points on the wing. They can be distributed differently depending on the mesh distribution type. The angle of attack represents the angle between the chord line of the wing and the vector representing the flow direction.

Two kinds of wing meshes are going to be studied, both symmetric with respect to the leading edge, $s = 0$: Elliptic mesh and Rectangular mesh, as shown in Figure 4.2.

Rectangular Mesh: A curvilinear rectangular mesh (see Figure 4.2-left) containing $N_s \times N_y = 100 \times 50 = 5,000$ points where all mesh points belong to the wing itself. They are uniformly distributed along both directions, chord-wise

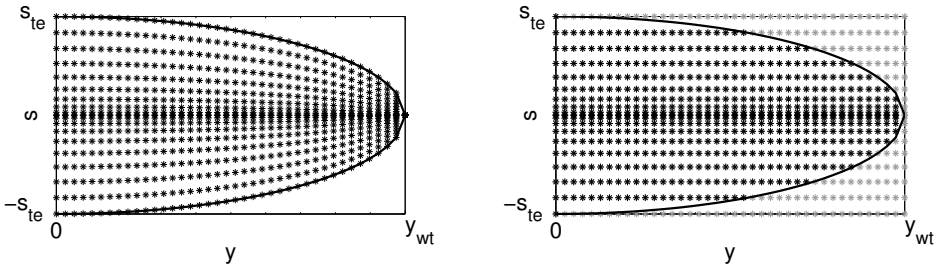

FIGURE 4.2: The rectangular (left) and elliptical (right) meshes for the elliptical wing model.

and span-wise. Note that this mesh requires considering an unnecessarily large number of mesh points near the wing tip.

Elliptic Mesh: The mesh is uniformly distributed along both spatial directions regardless of the physical limits of the wing. In other words, the wing corresponds to an elliptic region (see Figure 4.2-right) of a larger embedding structured mesh that exhibits the same number of points, $N_s \times N_y = 100 \times 50 = 5,000$, as the rectangular mesh. The elliptic region contains a smaller number of mesh points, namely 4,184 points within the wing.

In both meshes, the grid points are equispaced along the span coordinate y but, in order to account for the steep pressure gradient near the leading edge, grid points in the coordinate s (along the chord) are concentrated near the leading edge, $s = 0$, using a sinusoidal distribution. For illustration, a tensor slide representation at constant angle of attack, $\alpha = 4°$, for both the rectangular and elliptic meshes is given in Figure 4.3. It is important to highlight that the missing data region in the elliptic mesh is represented in this plot by a white grid, but in the following plots those points are omitted (but considered). Figure 4.3-right shows the irregular boundaries of the elliptic region containing C_p values on the wing, inside the larger rectangular mesh. Note that, in spite of having concen-

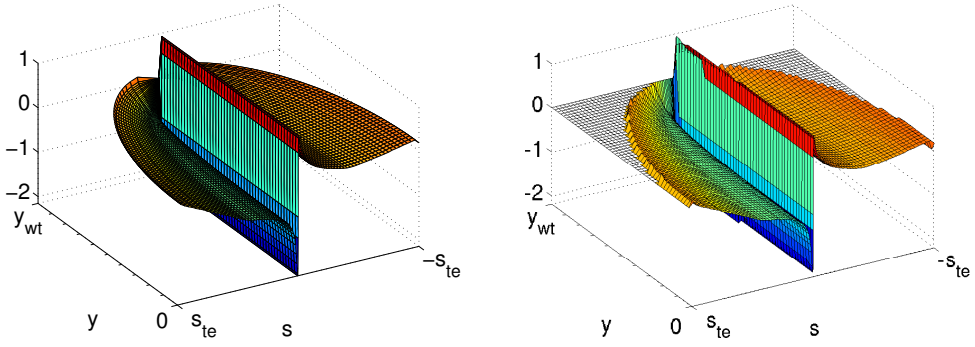

FIGURE 4.3: The pressure coefficient for $\alpha = 4°$ in the rectangular (left) and elliptic (right) meshes illustrated in Figure 4.2.

trated the mesh points near the leading edge, a very steep pressure jump still remains near the leading edge. This is one of the characteristics of typical aerodynamics database that challenges the developed method. In addition, there is the already mentioned three-dimensional artifact near the wing tip. These hinder gappy reconstruction, especially when gappy points are located either near the leading edge or near the wing tip.

The angle of attack is considered in the range $-2^\circ \leq \alpha \leq 6^\circ$, which is discretized considering 5 equispaced values. All these give two three-dimensional databases, one for the rectangular mesh, whose size is

$$N_s \times N_y \times N_\alpha = 100 \times 50 \times 5 = 25,000 \quad \text{points}, \tag{4.1}$$

and another of the same size for the augmented structured mesh containing the smaller (20,920 mesh points) elliptic region associated with the wing itself.

4.2 Theoretical database analysis

Two different mesh types have been analysed for assessing two different approaches in data collection and storage in experimental tests. The rectangular mesh is commonly used in numerical simulations, as CFD analysis, where special regions, as the wing tip or the leading edge, require higher density of cells to improve the convergence and/or precision of the solution. Instead, the elliptical mesh is typically chosen for experimental tests in wind tunnel. Here, the wing tip of the model is a very small region where placing a high number of sensors is physically impossible. Furthermore, characteristic patterns of these databases due to the aerodynamic nature of the data make them very demanding from the reconstruction point of view. The wing tip artifact involves redundancies that:

- Are very different to the aerodynamic redundancies in the remaining part of the wing.
- Are poorly defined because the correction affects a very small region near the wing tip.

These theoretical databases are free of noise and will be considered as exact baseline solutions. This aspect allows to better study the proposed methods performance without inherent errors in the original database that could interfere in the results. Thus, the performance of the noisy-gappy-HOSVD method will be assessed by adding random noise to an initially 'clean' data.

As detailed in the previous section, both databases can be written as three dimensional tensors of the same size (Eq. (4.1)) where randomly distributed or concentrated gappyness will be present. The incomplete and/or contaminated tensors reconstruction will be evaluated by measuring errors in terms of the RMSE and MaxE, defined in Eq. (2.14), in the same way as presented in §3.1

and §3.2. Let us recall that errors in the gappy and non-gappy regions of the database, the same as truncated HOSVD errors as a reference, will be considered. The errors $(\text{RMSE}^{\text{cont}}, \text{MaxE}^{\text{cont}})$ and $(\text{RMSE}^{\text{clean}}, \text{MaxE}^{\text{clean}})$, defined in Eqs. (3.9)-(3.10) will be also included when noisy databases are treated.

All error values are normalized with $(\max_{ijk}(A_{ijk}) - \min_{ijk}(A_{ijk}))$, the maximum and minimum values of the clean tensor A, respectively, in order to facilitate their interpretation in comparison with different databases.

Now, the methods proposed in chapter §3 can be directly applied in the rectangular mesh (see §4.2.1), but not in the elliptic mesh, in which the wing does not correspond to a structured mesh. This difficulty will be firstly solved in §4.2.2 and then, gappyness reconstruction for this mesh is analysed in §4.2.3. Furthermore, a new strategy is proposed in §4.2.4 in order to reconstruct gappyness dealing with portions of the database instead of the complete one.

4.2.1 Gappy reconstruction in the rectangular mesh

Let us now reconstruct gappy data in the rectangular mesh. As in §3.1 and §3.2, both randomly distributed and concentrated gappyness, for both the 'clean' database and a contaminated database, are considered in the rectangular mesh. The parameters of the gappy-HOSVD and noisy-gappy-HOSVD methods are the same as those used for the toy model database and the performance of the methods is also similar. These illustrate well the robustness of the methods.

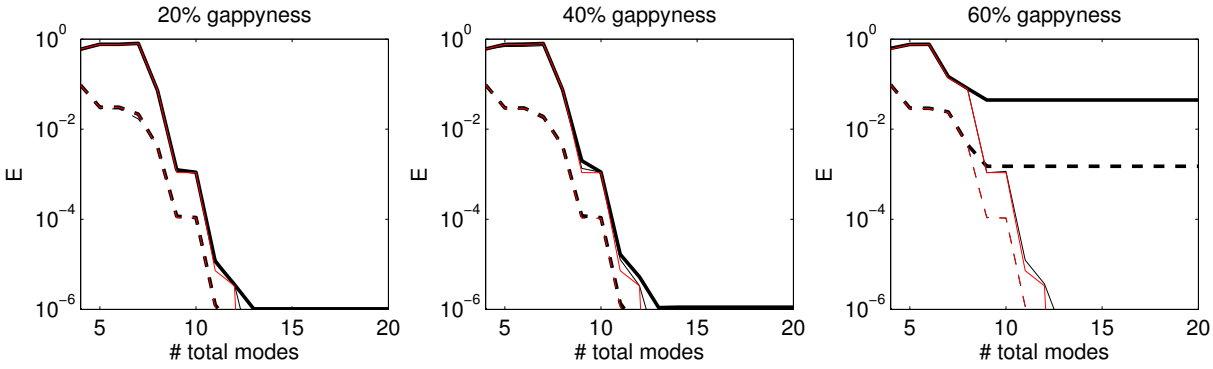

FIGURE 4.4: Application of the gappy-HOSVD method to the theoretical aerodynamic database in the rectangular mesh for randomly distributed gappyness with the indicated gapyness level, plotting RMSE (dashed lines) and MaxE (solid lines), as defined in Eq. (2.14), of the reconstructed solution in the gappy (thick-black lines), the non-gappy (thin-black lines) regions, and the result of applying truncated HOSVD to the clean database retaining the same numbers of modes (thin red lines).

For randomly distributed gappyness, Figure 4.4 shows that, as in Figure 3.1:

- The reconstructions are optimal (namely, the associated errors are similar to their counterparts for the clean database retaining the same numbers of modes) up to a certain number of retained modes.
- The errors steadily decrease as the number of retained modes increases.

The reconstruction is extremely good for 20% and even 40% gappyness levels, retaining only $m+p+q = 13$ modes, since MaxE and RMSE for both are smaller than 10^{-6} (see Figure 4.4). Thus, the counterpart of Figure 4.3-left obtained from the gappy reconstructions are plot-indistinguishable from the exact plot. Moreover, the reconstruction with 60% gappyness is still fairly good, as Figure 4.5 shows, where left plot represent the initial condition of the gappy tensor as specified in Eq. (3.1). In this case, the RMS and maximum errors retaining 10 modes are RMSE$= 1.5 \cdot 10^{-3}$ and MaxE$= 4.4 \cdot 10^{-2}$, respectively. This is remarkable for this gappyness level, taking into account that a significant part of the reconstructed data are located in the large-steep-gradient region near the leading edge.

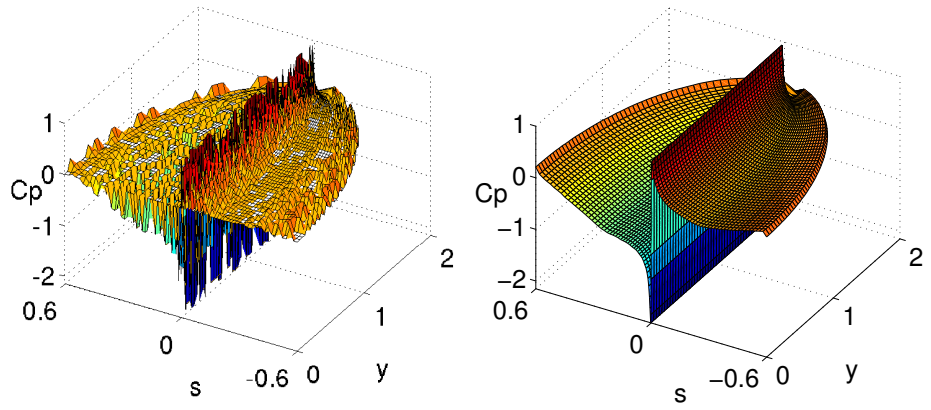

FIGURE 4.5: The pressure coefficient for $\alpha = 4°$ with 60% gappyness (left) and its reconstruction using $m + p + q = 10$ modes (right).

Concerning concentrated gappyness, the black-box gappy region

$$-s_{\text{te}} \leq s \leq s_{\text{te}}/4, \quad y_{\text{wt}}/3 \leq y \leq 2y_{\text{wt}}/3, \tag{4.2}$$

is a significant part of the wing (specifically, 23.4% of the database) and includes a significant part of the leading edge, where the pressure gradient is quite steep, as illustrated in Figure 4.6-top-left. Although its gappyness level is comparable to the first plot in Figure 4.4, the differences between both reconstruction error plots demonstrate how demanding the reconstruction of concentrated gappyness is. In spite of this, the gappy reconstruction is quite good. In particular, the reconstructed database retaining $m + p + q = 9$ modes exhibits a MaxE$= 6.4 \cdot 10^{-3}$, which makes the reconstructed database (Figure 4.6-top-right) plot-indistinguishable from its original counterpart.

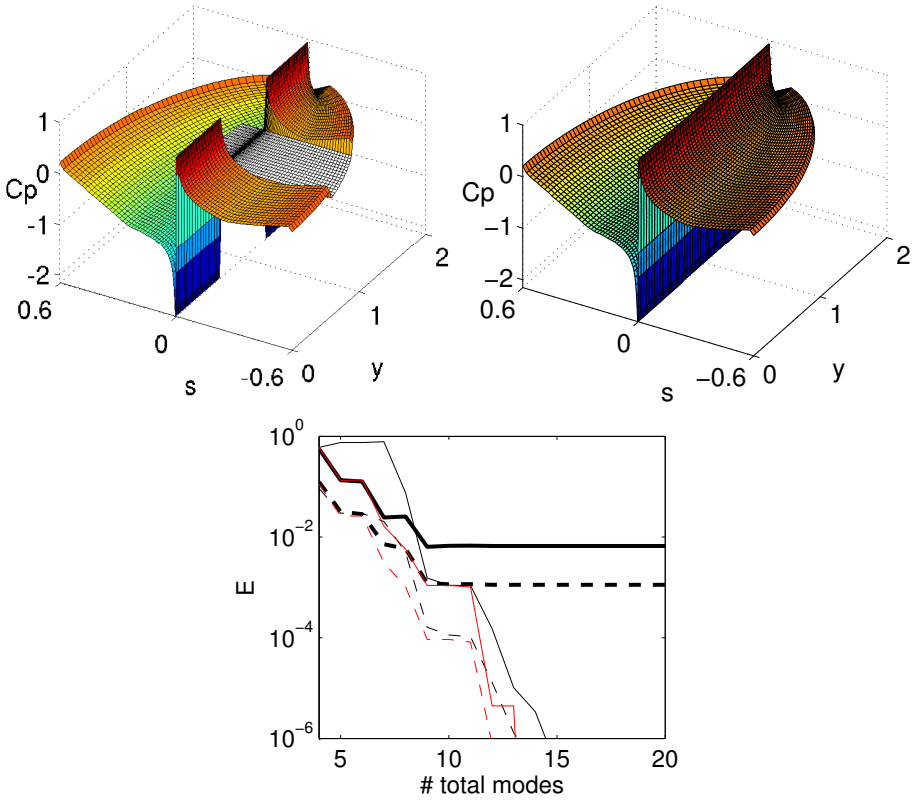

FIGURE 4.6: Counterpart of Figure 4.4 for the concentrated gappy database (bottom) illustrated in the top-left plot for $\alpha = 4°$; the reconstruction using $m+p+q = 9$ modes is given in the top-right plot.

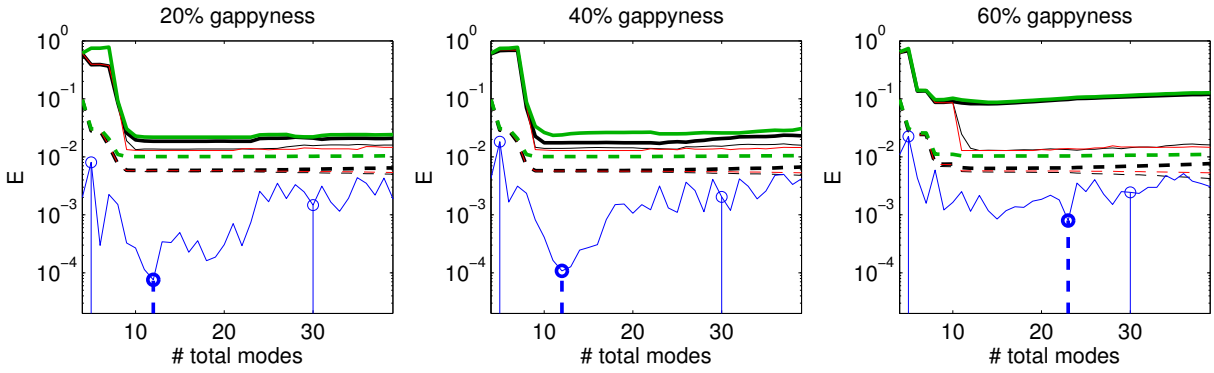

FIGURE 4.7: Counterpart of Figure 4.4 adding positive, uniformly distributed, noise of size 0.02 and applying the gappy-noisy-HOSVD method; the MSE estimate is plotted in solid blue line. As in Figure 3.7, the reconstruction errors in the gappy region resulting from comparison with the contaminated ($\text{RMSE}^{\text{cont}}$, $\text{MaxE}^{\text{cont}}$) and clean ($\text{RMSE}^{\text{clean}}$, $\text{MaxE}^{\text{clean}}$) databases are plotted with thick black and green lines, respectively.

Let us now add some random noise (positive, uniformly distributed noise, as

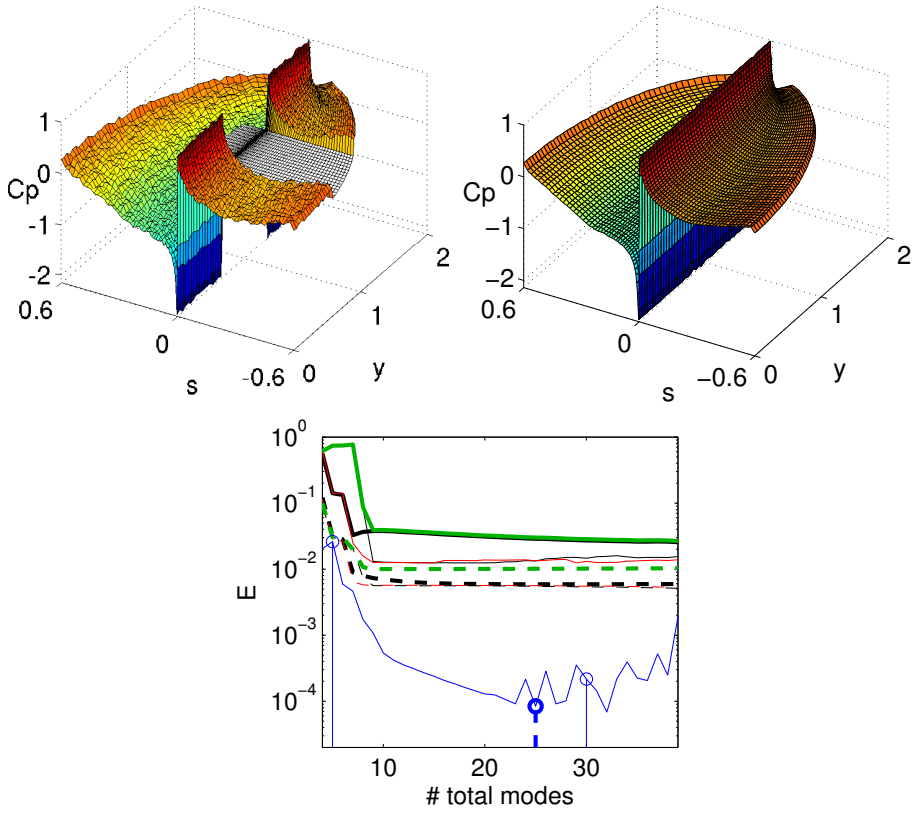

FIGURE 4.8: Counterparts of Figure 4.6 adding positive, uniformly distributed, noise of size 0.02 and applying the gappy-noisy-HOSVD method; the MSE estimate is plotted in solid blue line. As in Figure 3.9, the reconstruction errors in the gappy region resulting from comparison with the contaminated ($\text{RMSE}^{\text{cont}}, \text{MaxE}^{\text{cont}}$) and clean ($\text{RMSE}^{\text{clean}}, \text{MaxE}^{\text{clean}}$) databases are plotted with thick black and green lines, respectively.

in §3.2) of size 0.02, to the aerodynamic database and then reconstruct both randomly and concentrated gappyness. Applying the gappy-noisy-HOSVD method, with $N_{\text{min}} = 5$ and $N_{\text{max}} = 30$, the original database is reconstructed within the errors plotted in Figures 4.7 and 4.8.

As can be seen, the mode selection estimate MSE, defined in Eq. (3.7), makes a good selection of the optimal number of retained modes. Although the counterpart of Figure 3.2 is omitted, the reconstructed database with the selected numbers of modes is plot indistinguishable from the clean database for randomly distributed gappyness. Also, as in the toy model database, the errors calculated according to Eqs. (3.9) and (3.10) follow similar trends. For concentrated gappyness (Figure 4.8-top), the reconstruction is still reasonably good. Note that the selected number of modes (25) is larger than for randomly distributed gappyness, as expected. It should be noticed also, that this reconstruction involves extrapolation along the s-coordinate, which is more demanding as shown in §3.1.

Moreover, $\text{RMSE}^{\text{clean}}$ (Figures 4.7 and 4.8) is smaller than the added noise even for 60% randomly distributed gappyness; $\text{MaxE}^{\text{clean}}$ is comparable to the added noise for the corresponding 20%, 40%, and black-box gappyness cases. In other words, the method filters the added noise out. This is in spite of the large pressure gradient in a part of the gappy region and it should be remarkable that the added noise does not exhibit a zero mean.

4.2.2 Gappy completion of the incomplete elliptic mesh

The elliptic mesh, as defined in §4.1, can be considered as a real example of structured but non-rectangular mesh contained in a larger rectangular mesh, which can be treated as outlined at the end of §3.1. In other words, those points not pertaining to the wing in the larger rectangular mesh (gray points in Figure 4.2-right) are treated as gappy points, which are reconstructed using the gappy-HOSVD method described in §3.1. The HOSVD description limited to those points inside the elliptic region (which are precisely the points pertaining to the wing) shows a reasonably good behaviour, as seen in Figure 4.9. In par-

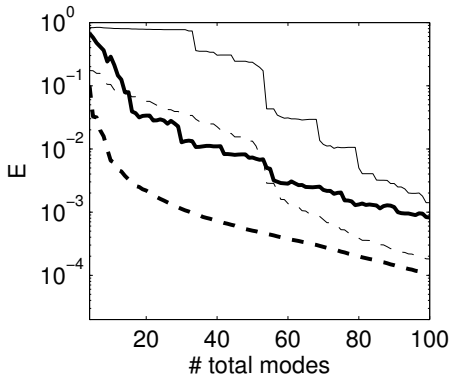

FIGURE 4.9: RMSE (dashed lines) and MaxE (solid lines) of the reconstructed solution in the interior region, via the gappy-HOSVD method (thick lines), and the standard truncated HOSVD in the larger rectangular mesh setting to zero the pressure coefficients at the unphysical points (thin lines).

ticular, retaining $m + p + q = 30$ modes yields the following errors in the wing: RMSE$= 1.1 \cdot 10^{-3}$ and MaxE$= 1.3 \cdot 10^{-2}$. It is interesting to remark that these errors are much smaller than their counterparts (thin lines plotted in Figure 4.9) obtained by applying truncated HOSVD to the database where the unphysical points (outside the elliptic domain) are set to $C_p = 0$. Setting to zero the unknown points is the most obvious way to treat this non-rectangular databases using HOSVD. In fact, it coincides with the initial reconstruction (3.1) used for the gappy-HOSVD method. But it leads to fairly bad results, as seen in Figure 4.9.

Summarizing, the method based on gappy-HOSVD yields a fairly good HOSVD description of the database in the non-rectangular, elliptic mesh. In

particular, the performance of the method is generally better than the alternative method mentioned above.

4.2.3 Gappy reconstruction in the elliptic mesh

Previous section addressed the problem about having unphysical points in an elliptic mesh-like database. Treating these particular points as a gappy region helps to the proposed methods to better reconstruct the points inside the elliptic domain. Now the reconstruction of randomly distributed and concentrated gappyness (as done in §4.2.1 for the rectangular mesh) in the elliptic mesh is addressed.

This study case is very demanding because in comparison with the rectangular mesh, it tries to reproduce the same physics but with less available information. The gappy-noisy-HOSVD method is proposed to be used in this case because the points outside the elliptic domain (more than 16% of the database) together with the induced gappyness make the error curves trend behave as in noisy databases. This feature can be observed in both Figure 4.10 and 4.11, since error curves in the gappy region decrease when the number of retained modes increases to then increase.

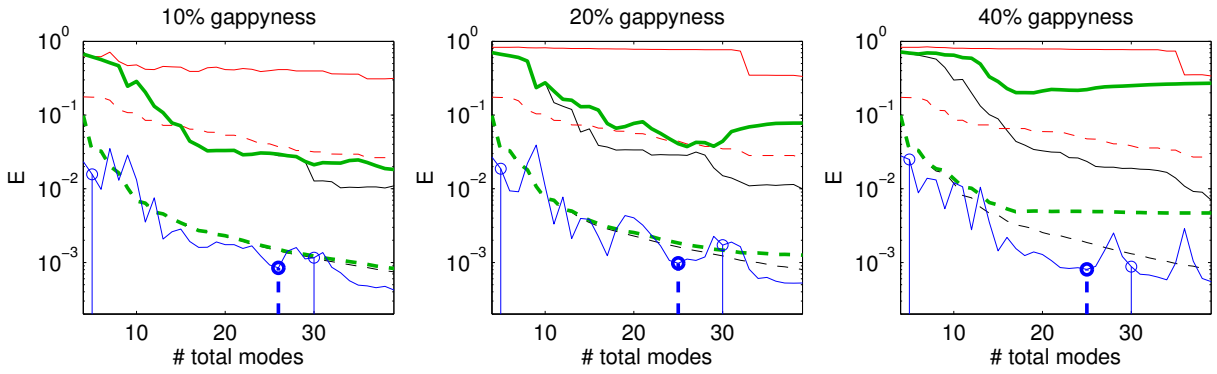

FIGURE 4.10: Counterpart of Figure 4.4 applying the gappy-noisy-HOSVD method for the elliptic mesh; the $\text{RMSE}^{\text{cont}}$, $\text{MaxE}^{\text{cont}}$ curves (giving the errors comparing with the contaminated database) are not plotted because the database is cosidered as 'clean', whithout added noise.

The good performance of the gappy-HOSVD description observed in the previous section is also confirmed in this case, where the truncated HOSVD method is not able to reduce the MaxE below 10^{-1} in the non gappy region while the gappy-noisy-HOSVD method does.

Gappyness reconstructions show higher errors and higher number of retained modes in comparison with their counterparts of the rectangular mesh case, as expected, due to the already mentioned difficulties of the elliptic mesh. It is interesting to notice that the indicated percentage of randomly distributed gappyness (in Figure 4.10) is applied to the elliptic region only, hence the points

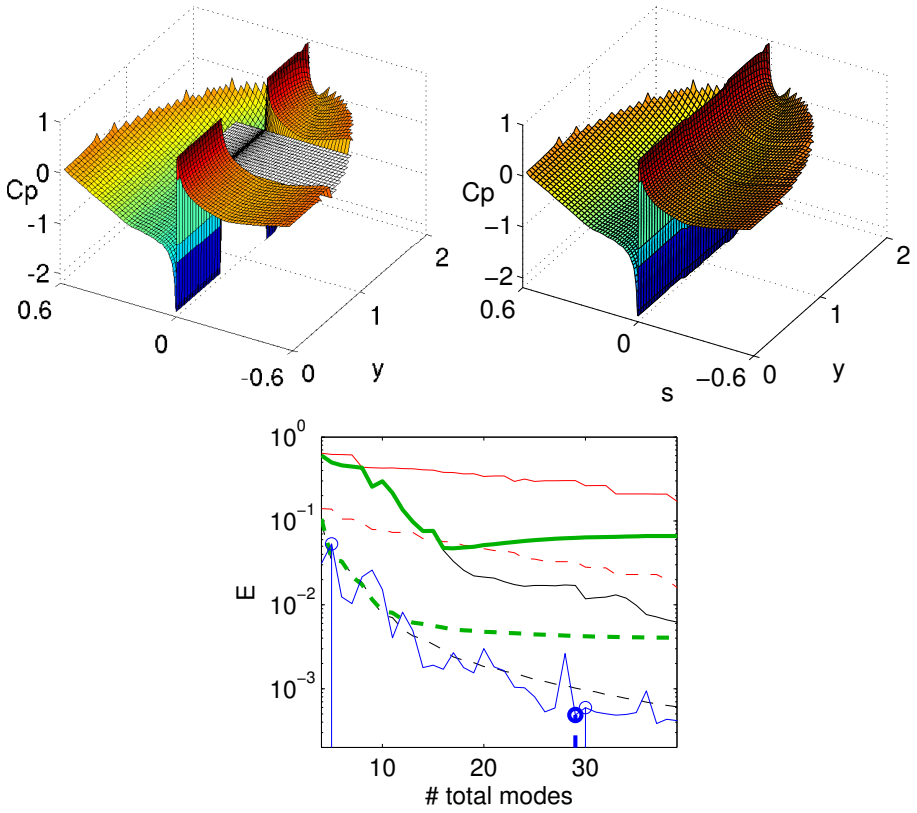

FIGURE 4.11: Counterpart of Figure 4.6 applying the gappy-noisy-HOSVD method for the elliptic mesh; the $\text{RMSE}^{\text{cont}}$, $\text{MaxE}^{\text{cont}}$ curves (giving the errros comparing with the contaminated database) are not plotted because the database is cosidered as 'clean', whithout added noise.

pertaining to the wing, yielding a percentage of points with clean information of 75%, 67% and 50% instead of 90%, 80% and 60%, respectively. It is obvious since the non physical points are not missing data. In the case of 10% random gappyness when $m + p + q = 26$ modes are retained, MaxE$= 3.0 \cdot 10^{-2}$, which is equal to the one obtained for the rectangular mesh database with 20% random gappyness and noise level of 0.02. As anticipated, the unphysical points introduce some noise in the reconstruction performance, especially in the 20% gappyness case. Using the gappy-noisy-HOSVD method clearly helps in the number of retained modes selection as the black-box gappyness reconstruction shows. In this case, the optimal number of retained modes are $m + p + q = 16$ modes, where MaxE$= 4.7 \cdot 10^{-2}$ and RMSE$= 5.3 \cdot 10^{-3}$. The MSE selection slightly reaches close values at $m + p + q = 29$ modes, where MaxE$= 6.3 \cdot 10^{-2}$ and RMSE$= 4.3 \cdot 10^{-3}$.

For concentrated gappyness the method works well also as Figure 4.11-top shows, where the gappy region reconstruction is little distinguishable from the original. This is remarkable because again, the method is reconstructing simultaneously not only the genuine gappy region, but also the unphysical gappy

region needed to complete the description in the larger rectangular mesh with less than 62% of total points as non-gappy region.

4.2.4 Gappy-HOSVD method application to divided tensors

Previous sections demonstrate that the gappy-HOSVD method can be applied to non-rectangular domains as well, expanding its use to a wider range of possible applications. Independently of the shape of the original database inside the tensor (rectangular domain or not), a different approach could be treating the database by splitting it. Dividing it into several regions, with a shape that is not longer restricted to be parallelepiped can be sometimes useful. Although redundancies are needed to achieve a better performance of the method, sometimes some of these redundancies can be lost in order to improve other aspects as avoiding abrupt regions. The gappy-HOSVD method (based on an extended version of standard SVD to higher dimensions) does not manage well such phenomena, instead such kind of regions makes the gappy-HOSVD-based methods not to work efficiently, inhibiting a good reconstruction. Splitting the domain into various smaller domains along the mentioned problematic regions could improve the method efficiency, as it reduces the amount of data to be processed in terms of space and time.

The reasons why a portion of the database should be used instead of the complete one can be very diverse and are not always related to the patterns of the database. Therefore, primarily a study on the toy model using different splitting schemes of the database is performed and then, a more representative case with the theoretical aerodynamic database is done.

The toy model database (shown in Figure 2.2) exhibits a wavy pattern in both x and y directions that dividing in smaller regions strictly including only one oscillation could complicate the reconstruction performance. With the aim of analysing the gappy-HOSVD method behaviour in this case, a concentrated gappyness square black-box type is considered, localized in an interior region defined by

$$0.34 < x < 0.54;\ 0.54 < y < 0.74;\ \forall(z) \tag{4.3}$$

Two kinds of division are proposed in Eq. (4.4), as Figure 4.12 shows. Top figures correspond with a middle slice of the original database in y-direction, in such a way that the middle oscillation in y-direction is isolated, named y-strip toy model database hereinafter. Bottom figures represent a middle slice in x-direction in order that same gappyness can be treated, named x-strip toy model database hereinafter.

$$\text{(a): } 0.3 < x < 0.6,\ \forall(y,z) \quad \text{(b): } 0.5 < y < 0.8,\ \forall(x,z) \tag{4.4}$$

For both x- and y-strip toy model databases (see Figure 4.13) note that around the gappyness only three database shells are known in y and x directions, respectively, in order to reconstruct such regions. As expected from conclusions

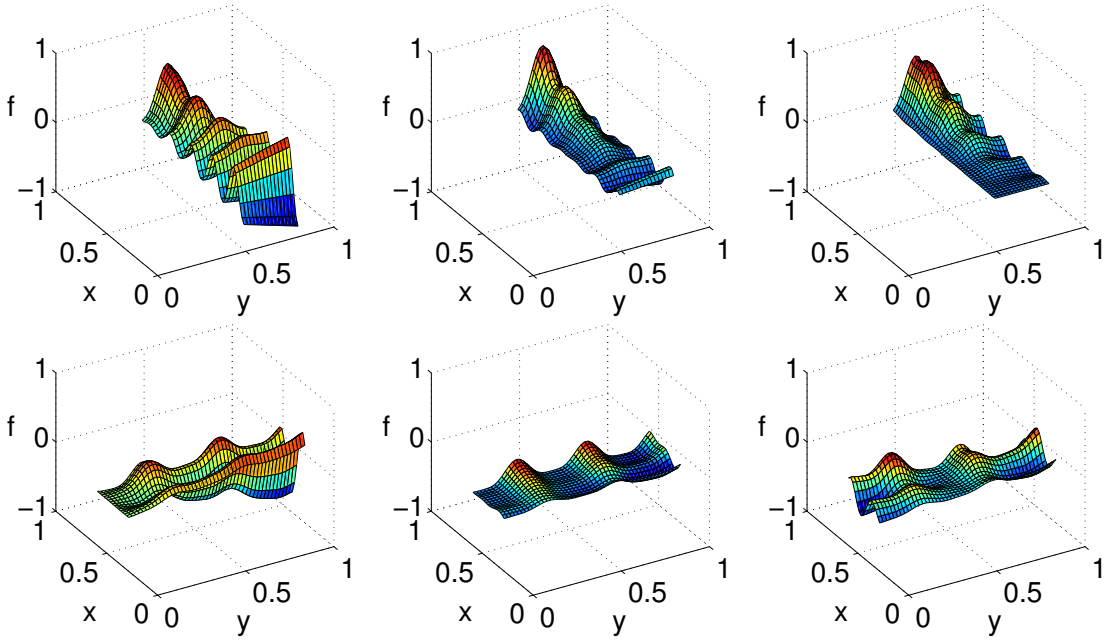

FIGURE 4.12: Counterparts of Figure 2.2-top where the original database has been sliced retaining only a middle oscillation in y-direction (top) and sliced in x-direction (bottom), acording to Eq. (4.4b) and Eq. (4.4a) respectively.

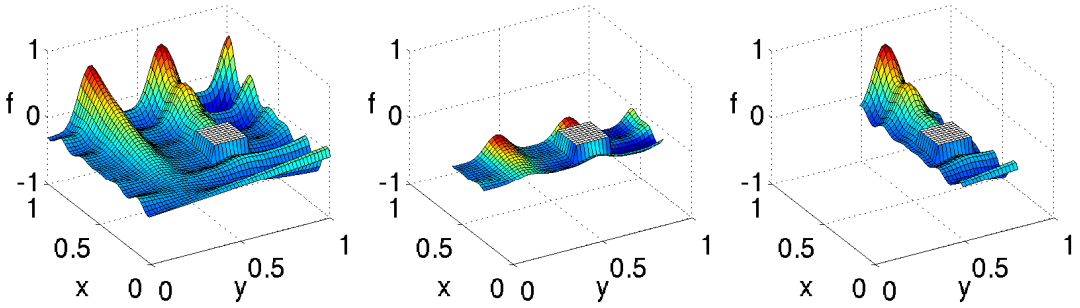

FIGURE 4.13: $z = 0.5$-cross-section of both x- and y-strip toy model databases, acording to Eq. (4.4a) (middle) and Eq. (4.4b) (right) respectively, with the gappy region defined in Eq. (4.3). For reference, the dense toy model with the same gappy region (left) is included.

in section §3, the gappyness defined in Eq. (4.3) is not highly demanding in terms of MaxE and RMSE for the dense toy model database. However, such conclusion is not obvious when same gappyness should be reconstructed in both strip toy model databases (see Figure 4.14). Despite being a portion of the dense toy model database, the x-strip toy model database exhibits a very efficient reconstruction since MaxE and RMSE in both gappy and non-gappy regions decay rapidly. It means that using a strip of 32% points from original, the gappy-HOSVD method advices a very high efficiency. On the other hand, the y-strip toy model database results are slightly worse, as the MaxE and RMSE in

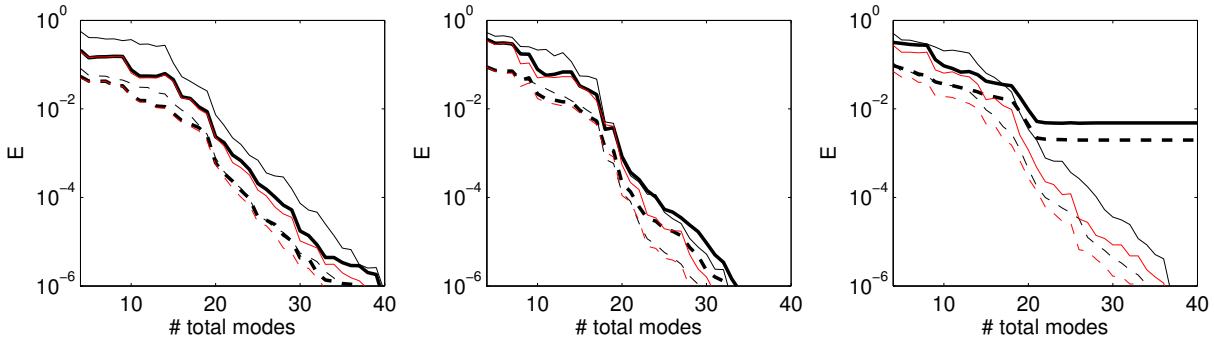

FIGURE 4.14: Counterparts of Figure 3.1 for the x- and y-strip toy model databases (middle and right plot, respectively) with the concentrated gappy region defined in Eq. (4.3). For reference, counterpart of Figure 3.1 for the dense toy model with the same gappy region (left plot) is included.

the gappy region achieve a minimum after $m + p + q = 21$ modes although it represents same 32% of the original database. It is due to the fact that this slice is on purpose selected including an isolated wave in y-direction and therefore, excludes highly relevant redundancies at the sides of the wave.

It is remarkable the good efficiency of the gappy-HOSVD method when using a portion of the complete database since MaxE in the gappy region remains always under 10^{-2} in both cases, in spite of loosing natural redundancies and correlation inside the database. Both analysis demonstrate that the strategy of dividing the original database could be profitable in terms of time and space computation since the TI is reduced from 1100 ($m + p + q = 19$ modes) to 821 ($m + p + q = 18$) and 1033 ($m + p + q = 20$ modes) in the dense, x-strip and y-strip toy model databases, respectively.

The comparison of the method efficiency when it is applied to both x- and y-strip toy model databases (see Figure 4.14) shows the importance of the strategy division since it is clearly related to the nature of database.

Next analysis tries to reproduce a real case applying a division strategy in the theoretical aerodynamic database with rectangular mesh taking in consideration the particular patterns of the database. As mentioned, a reason why splitting a database could be useful is that real aerodynamic databases may exhibit abrupt value changes in a concentrated region. This is indeed the case of the pressure coefficient distribution on a wing which exhibits high changes in pressure values in the proximity of the leading edge. Depending on the local density of mesh points when the geometrical domain is discretizated in this area, it could represent an abrupt change of patterns. As an example, this is the case of pressure measurement in experimental tests since placing high density of pressure gauges is physically impossible in so a small section of the chord as the leading edge of a wing model.

Figure 4.15 shows a section at α=4° of the original tensor and the three regions

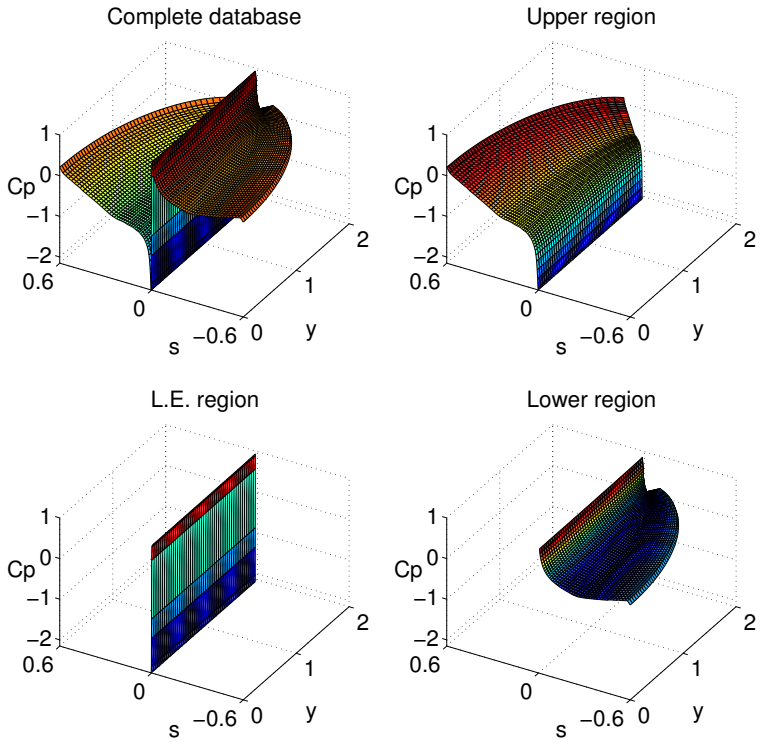

FIGURE 4.15: The pressure coefficient for $\alpha = 4°$ in the rectangular mesh database which has been split into three regions corresponding to the main wing regions: upper surface, lower surface and the leading edge.

in which it has been split into, corresponding to the main wing areas: upper surface (namely, upper region mesh), lower surface and the leading edge region. The division has been implemented according to the local slope of the C_p values in the s-coordinate direction. All resulting databases contain a piece of the original database whose patterns over all directions should be smooth. Those small databases can represent a non-parallelepiped database inside a larger rectangular mesh, as the case shown in section §4.2.2.

In this particular case, the L.E. region database contains few points in the s-coordinate direction which boundaries are also included in the upper and lower region databases. Nevertheless, the overlap rises open points regarding the division method and the gappyness position since various criteria of overlapping can be considered. This aspect is beyond the scope of this section and therefore is left for future work.

The division method does not affect the gappy-HOSVD method application since the possible missing points generated from a none rectangular division can be treated, as shown in previous sections.

A concentrated black-box type gappyness (Eq. (4.5)) has been considered to be in the upper surface region of the experiment, representing 30% of missing points in the rectangular mesh, as shown in Figure 4.16. However, the same

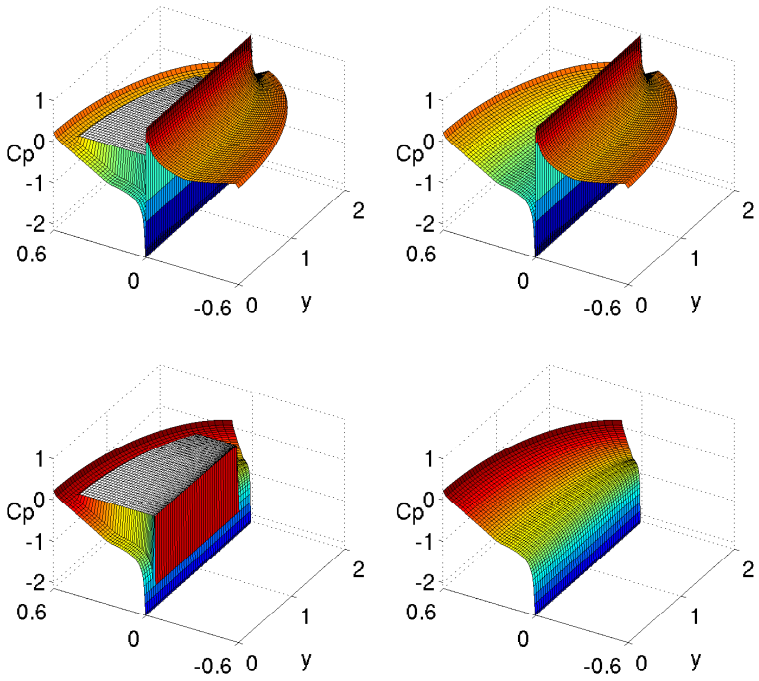

FIGURE 4.16: The concentrated gappy databases (left plots) of rectangular mesh (top) and upper region mesh (bottom), illustrated for α = 4°, with gappy region defined in Eq. (4.5) and their reconstructions (right plots) for $m+n+p=9$ modes and $m+n+p=7$ modes, respectively.

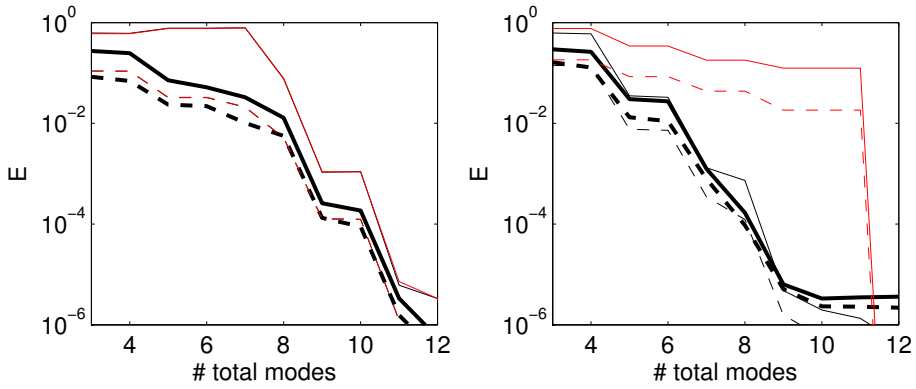

FIGURE 4.17: Counterparts of Figure 3.1 for the rectangular mesh (left) and the upper region mesh (right) with gappy region defined in Eq. (4.5).

gappyness represent 62% of points in the case of the upper region mesh. Despite this high level of gappyness and the loss of redundancies due to the strip along s-direction, the reconstruction accuracy is still good when applying the gappy-HOSVD method to the upper region mesh, as can be appreciated in Figure 4.16-bottom and in Figure 4.17-right. MaxE and RMSE decay spectrally in both the rectangular mesh and the upper region mesh remaining under 10^{-2} when $m+p+q>6$ modes and $m+p+q>8$ modes, respectively. These results

demonstrate the robustness of the gappy-HOSVD method since the divided tensor strategy joint with gappyness is a high demanding test case.

$$s_{\text{te}}/8 \leq s \leq 7s_{\text{te}}/8, \quad y_{\text{wt}}/8 \leq y \leq 7y_{\text{wt}}/8, \tag{4.5}$$

Futhermore, it should be noticed that Figure 4.17-right confirms again (as demonstrated in §4.2.2), that the gappy-HOSVD decomposition yields a fairly good HOSVD description of the stripped database instead of the truncated HOSVD. Previous conclusions about advantages of reconstructing concentrated missing data from a portion of the database are confirmed since same accuracy could be reached reducing the amount of information to be treated. The rectangular mesh requires retaining $m + p + q = 9$ modes and the upper region database $m + p + q = 7$ modes, where the truncated information, TI, (defined in Eq. (2.10)) to be managed is 520 and 214, respectively.

Summarizing, the developed method has been applied to a semi-empirical 'exact' model that reproduces most of the characteristics of a real aerodynamic database (mainly the discontinuity near the leading edge and non-rectangular meshes) with a good performance of the method. An extension considering the splitting of the tensor has also been tested that improves (but not substantially) the results at the cost of selecting manually the splitting criteria.

Chapter 5

Application to a CFD-generated aerodynamic database

The aim of this chapter is to achieve two specific closely related goals with two CFD-generated databases. First, the performance of the gappy-noisy-HOSVD method in relation to the aerodynamic databases size will be analysed using a precise database that represents a very good numerical approximation to the fluid field, where inherent errors and noise will be assumed to be negligible, in comparison with its coarser counterpart in section §4.2.1. In this previous section, the theoretical aerodynamic database was considered as exact and testing the performance of both the gappy-HOSVD and the gappy-noisy-HOSVD methods was possible. A similar situation will occur in CFD-generated databases when the CFD solver converges to the solution.

Furthermore, a second inaccurate database has been generated in order to check the proposed method performance when experimental noise and impure data are inherently present in the database and the 'exact' database is not known. This second database somehow resembles the behaviour that will be obtained when dealing with real experimental data.

5.1 CFD database generation

For illustration of a precise CFD-generated database, we consider the (incompressible) turbulent flow around a swept wing, with a Reynolds number based on the mean aerodynamic chord Re$= 4.4 \cdot 10^{-6}$. The wing planform is not flat, but exhibits some torsion. Specifically, the wing exhibits the following properties (see Figure 5.1): taper ratio=0.4, aspect ratio=9, washout angle=2°, quarter chord sweep angle=0°, and cross section aerofoil=NACA 64-210.

The CFD calculations were generated using OpenFOAM as CFD solver and Pointwise as mesh generator. Details are as follows:

- The superficial mesh was structured (quadrilateral), exhibiting 300×450 grid points in the chord/span directions, which were concentrated in the

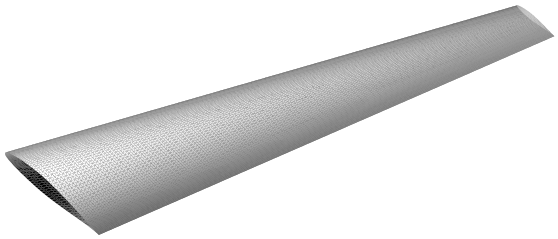

FIGURE 5.1: Counterpart of Figure 4.1 for the wing considered in the CFD-generated aerodynamic databases.

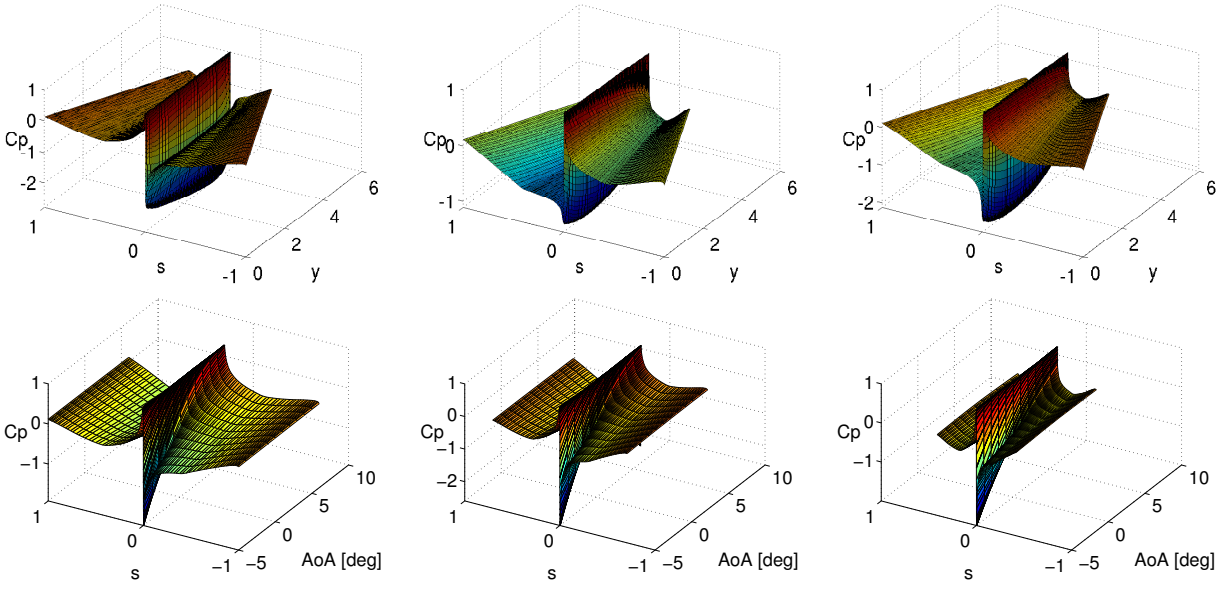

FIGURE 5.2: The pressure coefficient represented at constant α-value sections (top): $\alpha = -5°$, $\alpha = 4°$, and $\alpha = 6°$; and constant y-value sections (bottom): $y = 0$, $y = y_{wt}/2$, and $y = y_{wt}$.

leading edge zone. Thus, this mesh is somewhat similar to the rectangular mesh considered in Figure 4.2-left.

- Concerning the volumetric mesh, the boundary layer was extruded from the superficial mesh (using hexahedral cells) keeping the parameter y^+_{max} in the range $30 < y^+_{max} < 300$. The remaining part of the mesh was unstructured, using tetrahedral cells, keeping the mesh fairly fine in the proximity of the wing and gradually lowering the mesh-density towards the boundaries. The wake zone was also refined. The total number of cells was $6.3 \cdot 10^6$.
- This mesh can be considered as converged since the grid convergence index (GCI), based on the calculated lift force [50], is 0.26%.
- Since the flow is considered to be steady, the simpleFoam solver was used, with the Spalart-Allmaras turbulence model [51, 52] and Wall Functions to

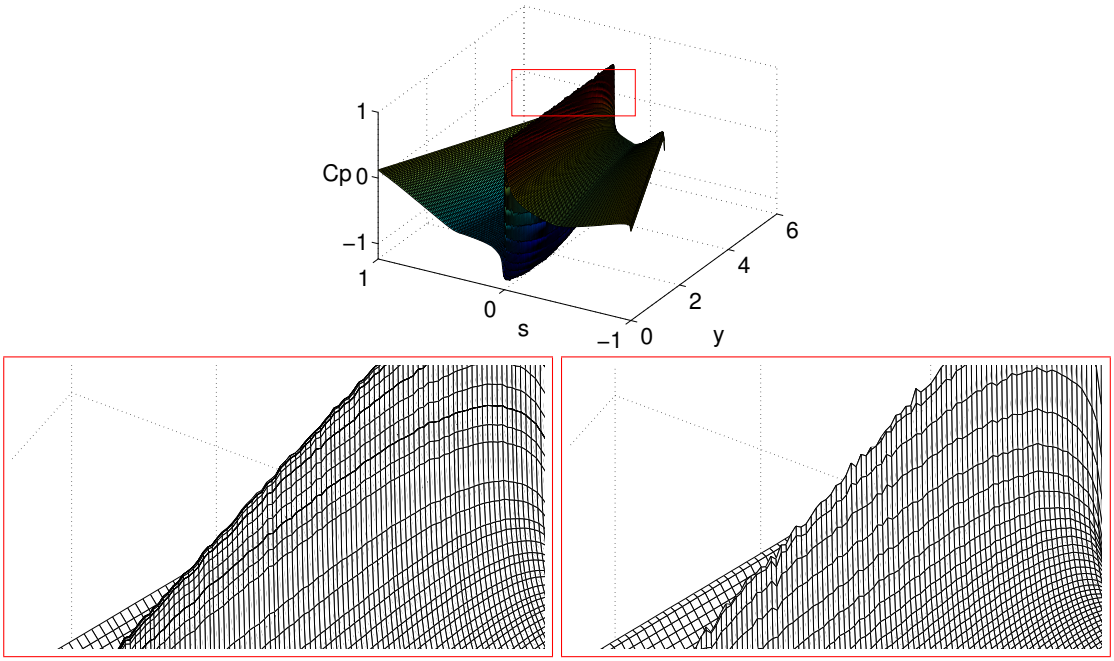

FIGURE 5.3: Zoom-in visualization near the leading edge region at $\alpha = 4°$ for both CFD-generated databases. Left plot corresponds to the precise well converged database and right plot corresponds to the inaccurate database.

approximate the flow near the wall. The artificial boundaries of the computational domain were placed far enough to simulate free flight. The spatial discretization was based on a second order scheme [53]. The computation was run until both the residuals and the aerodynamic coefficients converged.

Using the CFD solver, 12 equispaced values of the angle of attack are considered in the range $-5° \leq \alpha \leq 6°$. Taking into account the size of the superficial mesh, this gives a third order tensor of size $300 \times 450 \times 12$. This tensor is clearly more dense than its analogous in the chapter §4, as Figure 5.2 shows. Precisely, it allows a larger number of cells in the steepest variation region and therefore, a better performance of the proposed method could be expected.

However, the theoretical database considered in the chapter §4 and the CFD database considered above (in the following, referred to as *precise CFD-generated database*) are not a realistic situation keeping in mind industrial databases, which contain impure information. Databases obtained from either industrial CFD or wind tunnel tests usually exhibit several artifacts resulting from lack of precision and experimental errors, respectively. Furthermore, the 'exact' databases associated with physical phenomena are not known.

With these in mind, a new aerodynamic database is constructed. The wing is the same as in the previous case but an inaccurate geometric model has been used on purpose. The details of the OpenFOAM CFD solver are also the same

as before, except that (i) the kOmegaSST turbulence model is used (instead of Spalart-Allmaras), (ii) the size of the superficial mesh (coarser) is now 300×225, and (iii) the parameter y^+_{max} for extruding from the superficial mesh is taken as $y^+_{max} < 50$. In addition, a larger range of the angle of attack is considered, namely $-3° \leq \alpha \leq 10°$, where 14 equispaced values are selected, which recalling the size of the superficial mesh gives a three-dimensional database of size $300 \times 225 \times 14$.

Again the convergence of the analysis is ensured by both the residuals and the aerodynamic coefficients but the CFD implementation is done so as to simulate experimental errors that can occur if the physical model has some imperfections. In particular, these errors will be intentionally concentrated near the leading edge, where the abrupt change of the pressure coefficient in such a small region is defect prone. To that purpose, the so called *stagnation point anomaly*, characteristic drawback of the two-equation turbulence models [54] (i.e. the kOmegaSST used here), has been deliberately triggered in this database in order to obtained the desired behaviour without causing the divergence of the numerical solution. This effect can be observed in Figure 5.3 as a rippled behaviour over the stagnation points line ($C_p = 1$) of this *inaccurate CFD-generated database* (right zoom-in figure) in comparison with the almost perfect line of the precise CFD-generated database (left zoom-in figure). Furthermore, changes (ii) and (iii) provide the required inaccurate solution of the flow field due to a poorly constructed surface mesh. High aspect ratio cells are used in the superficial mesh since the discretization in y-direction is halved.

5.2 CFD database analysis

Two different CFD-generated databases have been introduced, each of one is going to be treated in the following using the gappy-noisy-HOSVD method. As in §4, all error values are normalized with $(\max_{ijk}(A_{ijk}) - \min_{ijk}(A_{ijk}))$, where A corresponds to the original tensor.

The precise CFD-generated database will be considered as the exact database (similarly to §4.2.1), although it is still inaccurate due to (small but non zero) discretization errors. Those errors are considered negligible since they do not complicate the gappy reconstruction. To emphasize robustness of the algorithm, the tunable parameters of the method are the same as those in the last sections. Moreover, as in the theoretical database, the tensor is contaminated with an uniformly distributed random error of size 0.02.

The application of the gappy-noisy-HOSVD method to the precise CFD-generated database is illustrated in Figures 5.4 and 5.5. As can be seen in both figures, the performance of the method is quite similar to that illustrated in Figures 4.7 and 4.8 for the theoretical database. In fact, the optimal numbers of modes are larger in the present case and the reconstruction errors smaller. Again, the mode selection estimate MSE defined in Eq. (3.7), plotted with solid

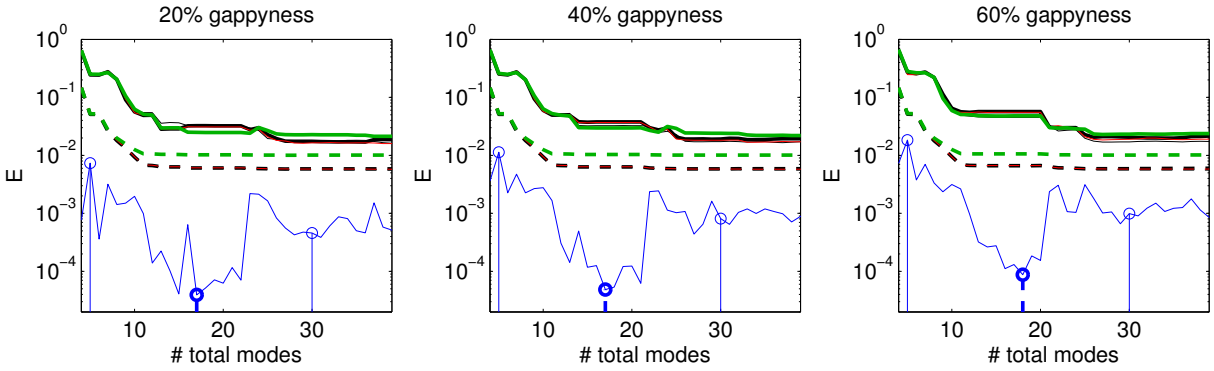

FIGURE 5.4: Counterpart of Figure 4.7 for the precise CFD-generated database.

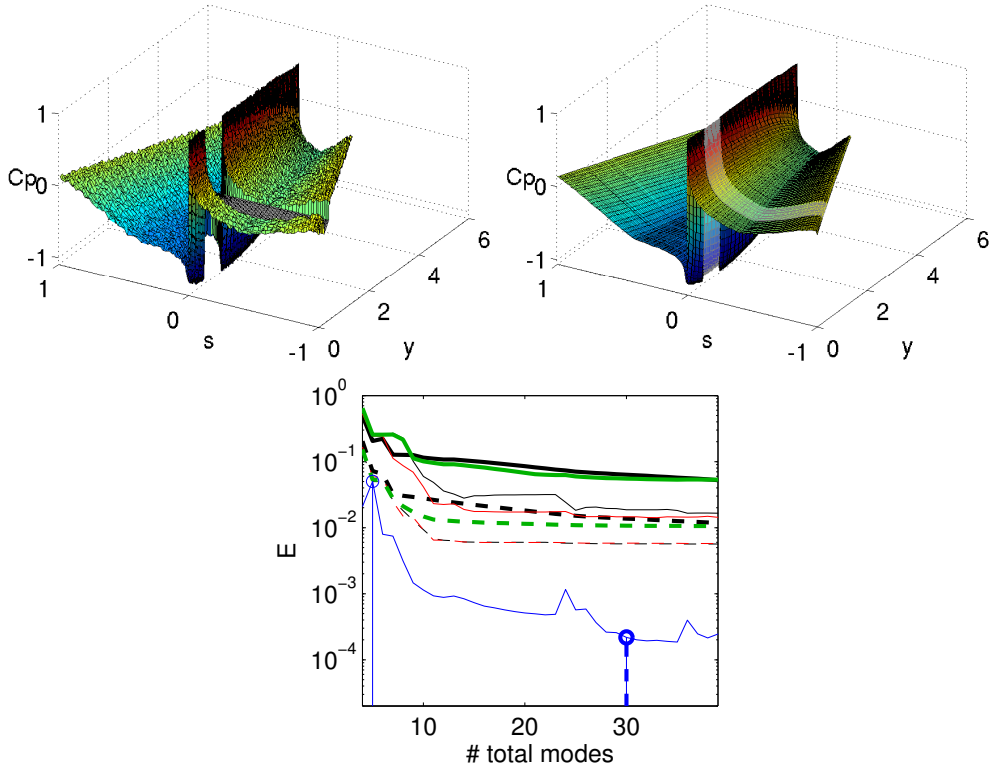

FIGURE 5.5: Counterpart of Figure 4.8 for the precise CFD-generated database.

blue lines, selects well the convenient number of modes. And $\text{RMSE}^{\text{clean}}$ is smaller than the size of the added error (of size 0.02), meaning that the method somehow filters the added error out. For random gappyness, the maximum error is comparable to the added error, while for concentrated gappyness, the maximum error is somewhat larger.

Considering the second CFD-generated database, as described in section §5.1, the geometric model used is inaccurate on purpose to simulate experimental errors that can occur if the physical model has some imperfections. Because of

these, the CFD results of this database are accurate only within 10^{-2} RMS errors and, in fact, localized errors of the order of 0.1 are present near the leading edge and the wing tip. In other words, the obtained database already contain significant errors, meaning that it is the gappy-noisy-HOSVD method that must be used to reconstruct gappy data. The performance of the proposed method for this new inaccurate database is given in Figures 5.6 and 5.7. As in the previous applications, to emphasize robustness, the tunable parameters of the method are the same as those in the former sections.

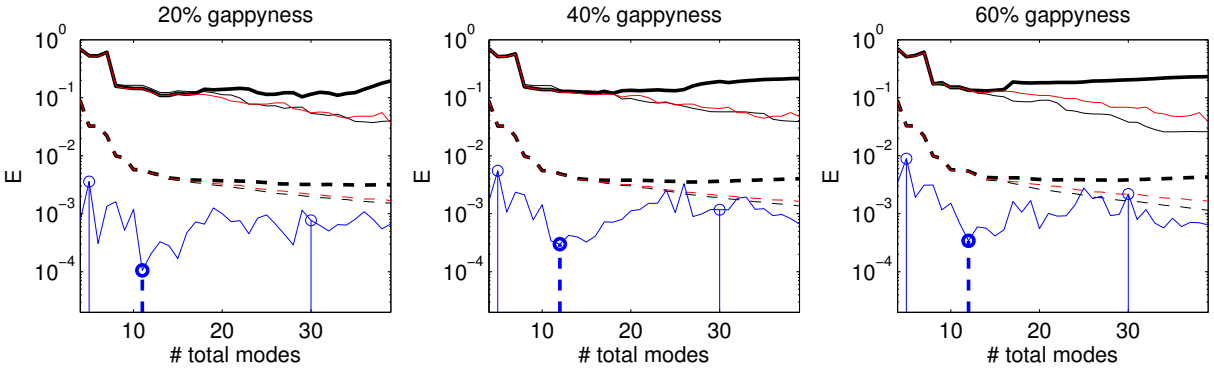

FIGURE 5.6: Counterpart of Figure 5.4 without added noise for the inaccurate CFD aerodynamic database. Note that, unlike in Figure 5.4, the green lines (giving the errors comparing with the clean database) are no longer plotted because the clean database is not known.

As can be seen in both figures, the plots of the RMSE and MaxE errors (calculated using Eq. (3.9)) vs. the selected number of modes is not monotone. Instead, the errors first decrease and then increase (as in Figures 4.7 and 4.8) suggesting that, as anticipated, truncation errors behave as noise. Note that:

- As in the previous applications of the method, the mode selection estimate MSE defined in Eq. (3.7), plotted with solid blue line in Figures 5.6 and 5.7, selects well the convenient number of modes.
- The maximum error MaxE is of the order of 0.1 in all cases, which is consistent with the discretization errors that are present in the database, as anticipated above. The RMS errors, instead, are smaller than the intrinsic noise, which, as in the previous applications of the method, means that the method filters these errors out. Particularly, it can be seen that the ripple effect observed in the Figure 5.3-right has been smoothly filtered for the reconstructed database, shown in Figure 5.8.
- For randomly located gappyness, the MaxE and RMSE at the selected numbers of modes are similar for the three levels of gappyness, which suggests that these errors mainly result from the truncation errors. In other words, the reconstructed database is seemingly better than the original database. This is consistent with what happened in everyone of the three databases, that are the toy model, the theoretical and precise CFD

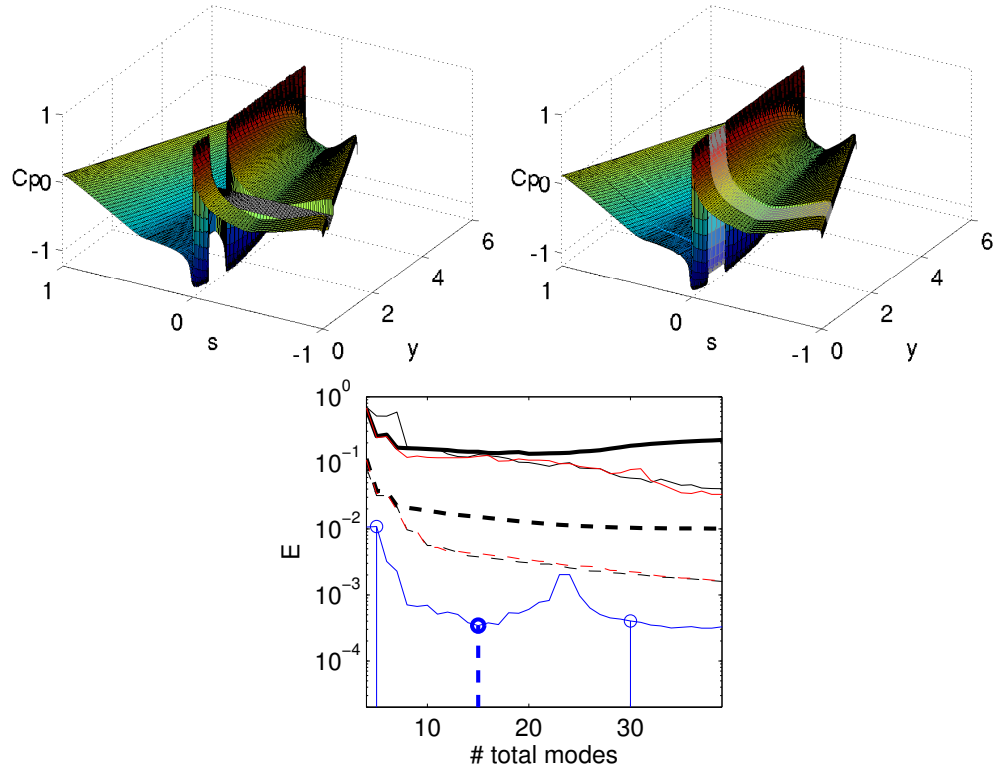

FIGURE 5.7: Counterpart of Figure 5.5 without added noise for the inaccurate CFD aerodynamic database. Note that, unlike in Figure 5.5, the green lines (giving the errors comparing with the clean database) are no longer plotted because the clean database is not known.

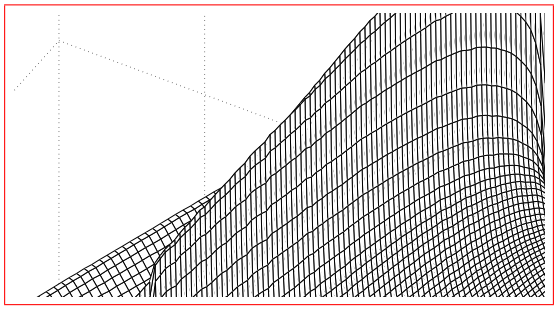

FIGURE 5.8: Counterpart of Figure 5.3 for the reconstructed inaccurate CFD-generated database. Inherent noise around the leading edge region observed in Figure 5.3-right has been almost completely filtered out.

aerodynamic ones, considered in Figures 3.7, 4.7 and 5.4, respectively, where noise was successfully filtered out by the method. Here, it is the (unknown) discretization errors (which act as noise) that are filtered out by the method.

- Unlike the toy model, the theoretical and the precise CFD aerodynamic

databases considered previously, this CFD-generated database produces error distributions for concentrated black-box gappyness (see Figure 5.7-bottom) that are quite similar to those resulting from random gappyness (see Figure 5.6). This confirms the previous conclusion that errors mainly result from the unknown discretization errors due to the CFD solver, and is consistent with the fact that the original database presents some wrinkle behaviour near the leading edge that is absent in the reconstructed database (see Figure 5.7-top).

Chapter 6

Application to an experimental database

Wind tunnel test campaigns produce rather big databases where several parameters come into play. It can be translated into high order tensors where measuring every test point involves a high cost. Budget and time limitations force to reduce the total number of tested points with respect to the total number of points of the high order tensor (affected by the curse of dimensionality), creating gappy regions inside this tensor. Independently of the inherent scientific interest these unknown points may have, the present Thesis demonstrated that filling these gappy points can help into the treatment of the complete tensor.

This chapter aims to apply the gappy-noisy-HOSVD method to an available aerodynamic experimental database which includes a high percentage of unknown points, with the intention of guessing the missing data in a way that they are well correlated with the known points of the database. In order to validate the method performance in this application, where the missing data is not known at all, a numerical approximation database has been calculated for reference.

6.1 Experimental testing campaign and database generation

A wind tunnel test campaign done in 2013 in the facilities of Instituto Tecnológico y de Energias Renovables (ITER) in Tenerife, Spain, collected experimental data for a box-wing configuration UAV design. The campaign was focused on studying the influence of main geometrical parameters that define the design on the aerodynamic coefficients and its overall efficiency.

The wind tunnel model was built as a modular fuselage where several variations of the lifting surfaces could be embedded (see Figure 6.1). In particular, five parameters define the degrees of freedom of the tested model which are detailed in Table 6.1. Main specifications of the wind tunnel model were: 2.25m length, 1.6m span and 0.2m wing chord, as shown in Figure 6.2. The nose of the

Parameters	Symbol	Values
Relative longitudinal separation of wings, adimensionalized by the chord and measured taking the leading edge of the root chord as reference.	L/c	[2, 3]
Relative vertical separation between wings, adimensionalized by the span and measured taking the leading edge of the root chord as reference.	G/b	[0.06, 0.085]
L.E. forward sweep angle of the front wing.	SW_{fr}	[0°, 23°, 40°]
L.E. backward sweep angle of the rear wing.	SW_{re}	[0°, 23°, 40°]
Dihedral angle of the rear wing, referred to as dihedral angle of both wings, since the anhedral angle of the front wing is forced to be equal in magnitude as the mentioned dihedral angle.	Dih	[0, 2.5°]

TABLE 6.1: Geometrical parameters of the box-wing configuration and their ranges studied during the test campaign.

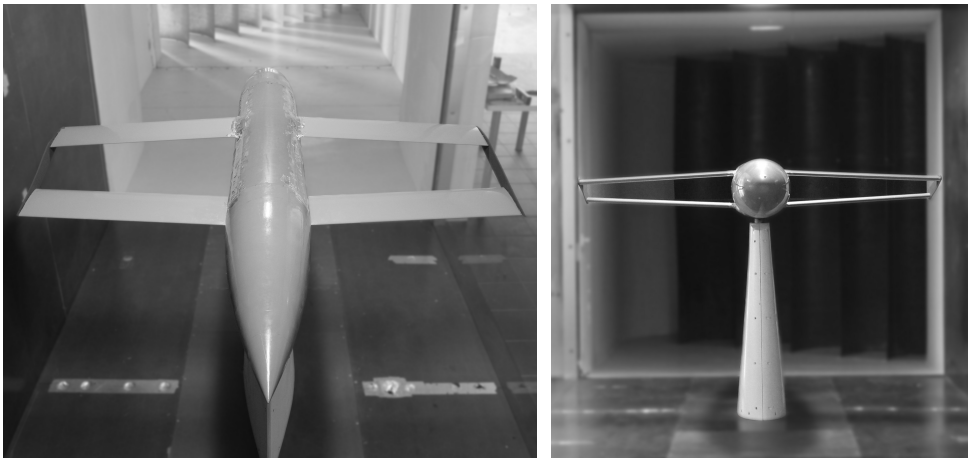

FIGURE 6.1: Top and frontal pictures of the wind tunnel model installed in the test chamber of the ITER Wind Tunnel.

fuselage is an ellipsoid of revolution connected with a cylindrical central section and the tail is designed with a circular arc as contour line. All lifting surfaces have an airfoil=NACA 0012, taper ratio=1, aspect ratio=8 and zero twist.

The complete map of variations from the baseline configuration to be studied consists on 72 combinations, as deduced from Table 6.1 but the limited resources allocated required to reduce the total number of tests to 27. All the configurations have been tested at two Reynolds numbers ($4.5 \cdot 10^5$ and $5.5 \cdot 10^5$) and a range of 21 angles of attack ($-19° \leq \alpha \leq 21°$).

The ITER wind tunnel facility is characterized by a test chamber section of 2.0×2.0m and a flow quality of 0.5% turbulence level. The aerodynamic forces

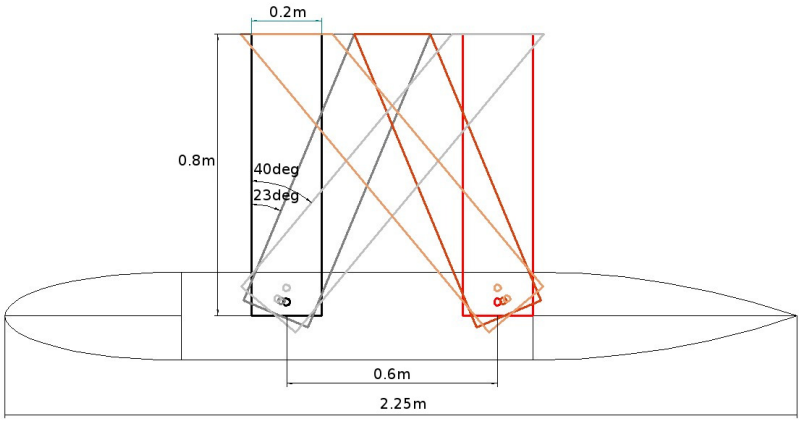

FIGURE 6.2: Wind tunnel model outline where several positions of sweep angle in both wings are shown.

and moments were measured by an external balance where the model is directly attached. Standard corrections for experimental test in wind tunnel, including as blockage, buoyancy, downwash and support corrections, have been applied to the complete experimental database, particularized for the specific characteristics of the tested model and the ITER Wind Tunnel.

Summarizing, the experimental database consists on the lift and drag coefficients (C_L and C_D, respectively) measured at a range of angles of attack and two different Reynolds numbers, of 27 variants of the box-wing configuration. Each database can be treated as a tensor of order 7 (see Eq. (6.2)) whose size is

$$\begin{aligned} N_{L/c} \times N_{G/b} \times N_{SW_{fr}} \times N_{SW_{re}} \times N_{Dih} \times N_{Re} \times N_{\alpha} = \\ = 2 \times 2 \times 3 \times 3 \times 2 \times 2 \times 21 = 3,024 \text{ points,} \end{aligned} \tag{6.1}$$

including many gappy points (62.5% of the total points). The gappyness can be considered as black-box because the non tested configurations represent a complete gappyness in the dimensions of both Reynolds number and angle of attack. The following data is considered:

$$\begin{aligned} C_L(L/c, G/b, SW_{fr}, SW_{re}, Dih, Re, \alpha) \\ C_D(L/c, G/b, SW_{fr}, SW_{re}, Dih, Re, \alpha) \end{aligned} \tag{6.2}$$

The 27 tested configurations were selected from the total 72 with the unique criterion of simplifying the testing procedure in order to minimize the required time for configuration changes, so no special interest was put in optimizing the parameter space exploration. It should be noticed that this fact will directly influence the goodness of the reconstruction results when the gappy-noisy-HOSVD method will be applied.

6.2 Numerical database generation

The aerodynamic experimental databases detailed in §6.1 are incomplete and make difficult to have a reference for assessing the accuracy of the reconstruction. For this reason, equivalent numerical databases has been calculated as an approximation of the complete experimental database space in order to evaluate the gappyness reconstruction performance. For this purpose, an in-house aerodynamic tool has been used able to quickly analyse parametrically variations of the box-wing configuration giving preliminary aerodynamic results from both aerodynamic methods Vortex Lattice and boundary layer study.

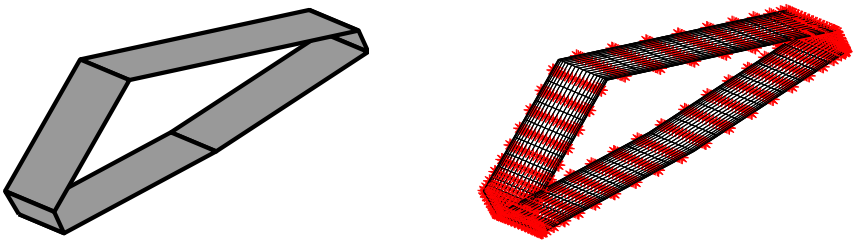

FIGURE 6.3: Box-wing configuration design sketched only as lifting surfaces on the left hand-side figure. Right hand-side figure represents the lifting surfaces division up into a lattice of quadrilateral panels, where the control points are highlighted in red.

Given both the set of geometric parameters (as defined in Table 6.1) and required flight conditions, this tool iteratively calculates each point in a few seconds. As output, lift and drag coefficients are given, similarly to the experimental ones defined in Eq. (6.2). The process consists of three steps: first, the geometry is discretized into a lattice of quadrilateral panels (see Figure 6.3), sufficiently small in order to have several panels in the tip surfaces. Then, the Vortex Lattice method step is applied where main aerodynamic performance is calculated based on the classical VLM method. Lastly, the boundary layer analysis method updates previous results adding viscous effects.

The numerical database content covers the whole map of 72 configurations analysed at a range of nine angles of attack ($-7° \leq \alpha \leq 9°$) and two Reynolds numbers ($4.5 \cdot 10^5$ and $5.5 \cdot 10^5$). Sumarizing, both C_L and C_D can be treated as tensors of order 7 whose size is $N_{L/c} \times N_{G/b} \times N_{SW_{fr}} \times N_{SW_{re}} \times N_{Dih} \times N_{Re} \times N_\alpha = 2 \times 2 \times 3 \times 3 \times 2 \times 2 \times 9 = 1,296$ points.

6.3 Experimental database analysis

Firstly, some differences between both experimental and numerical databases due to well-known limitations of Vortex Lattice methods should be noticed:

- The aerodynamic tool includes viscous analysis only for the C_D calculation, and therefore C_L shows no dependence on Reynolds number.

- As they are based on the potential theory, these methods cannot predict stall. This fact directly affects both C_L and C_D at higher angles of attack (in terms of absolute values), as flow separation cannot be predicted by this method.
- Wind tunnel tests were based on the analysis of the box-wing configuration with fuselage unlike numerical database which only consists of the lifting surfaces aerodynamics.

The limited validity range of the aerodynamic tool (with the intention of avoiding non-linear effects) constrained to reduce the calculated range of angles of attack for the numerical database in comparison with the experimental one. Furthermore, a fuselage correction has been applied to the experimental data. Thus, the corrected experimental database includes C_L and C_D values generated by the isolated wings as a whole.

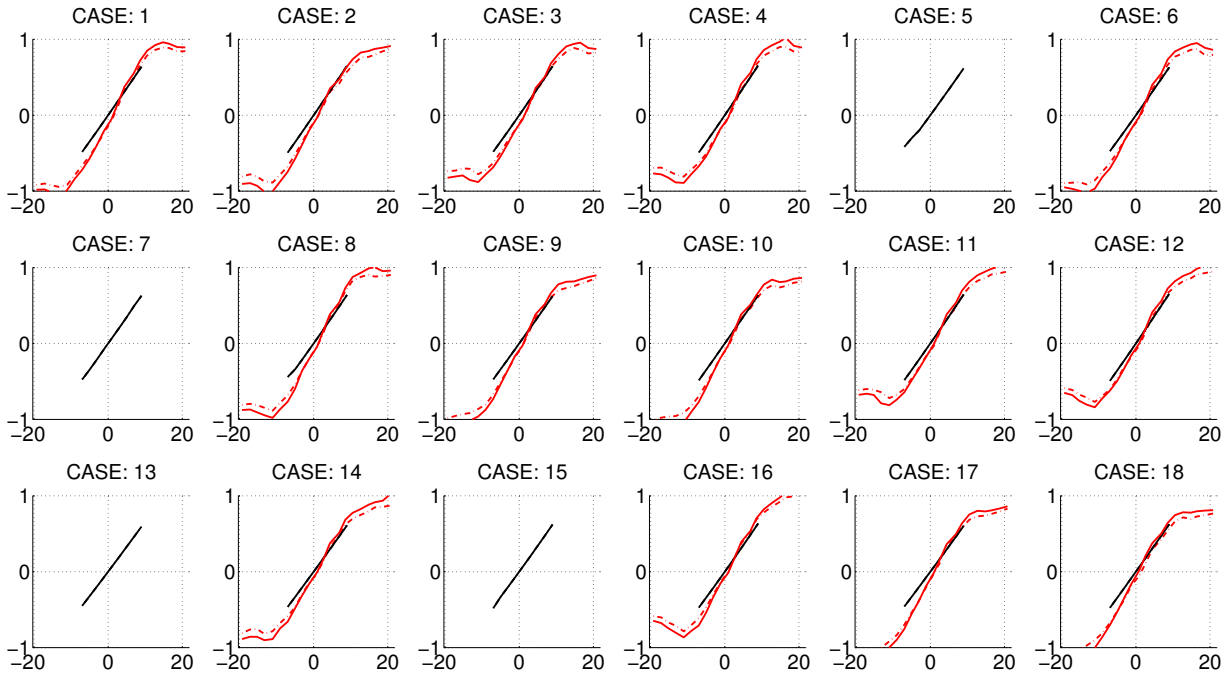

FIGURE 6.4: Lift coefficient versus α for both Reynolds numbers $4.5 \cdot 10^5$ (dot-dashed lines) and $5.5 \cdot 10^5$ (solid lines). Red lines represent the experimental database whereas black lines represent numerical database (Reynolds number indepent). Sample with only 18 variants detailed in Table B.1, Appendix B, of box-wing configuration out of the total 72.

Although the numerical method does not solve Navier-Stokes equations but uses potential flow instead, the coherence of both databases is demonstrated by the very good agreement between them shown in Figure 6.4 and Figure 6.5. The numerical method is only valid in the linear range of the flight conditions, in this case $-7° \leq \alpha \leq 9°$, where errors between both databases can be evaluated. Global differences in C_L are: MaxE=$3.1 \cdot 10^{-1}$ and RMSE=$9.2 \cdot 10^{-2}$ and

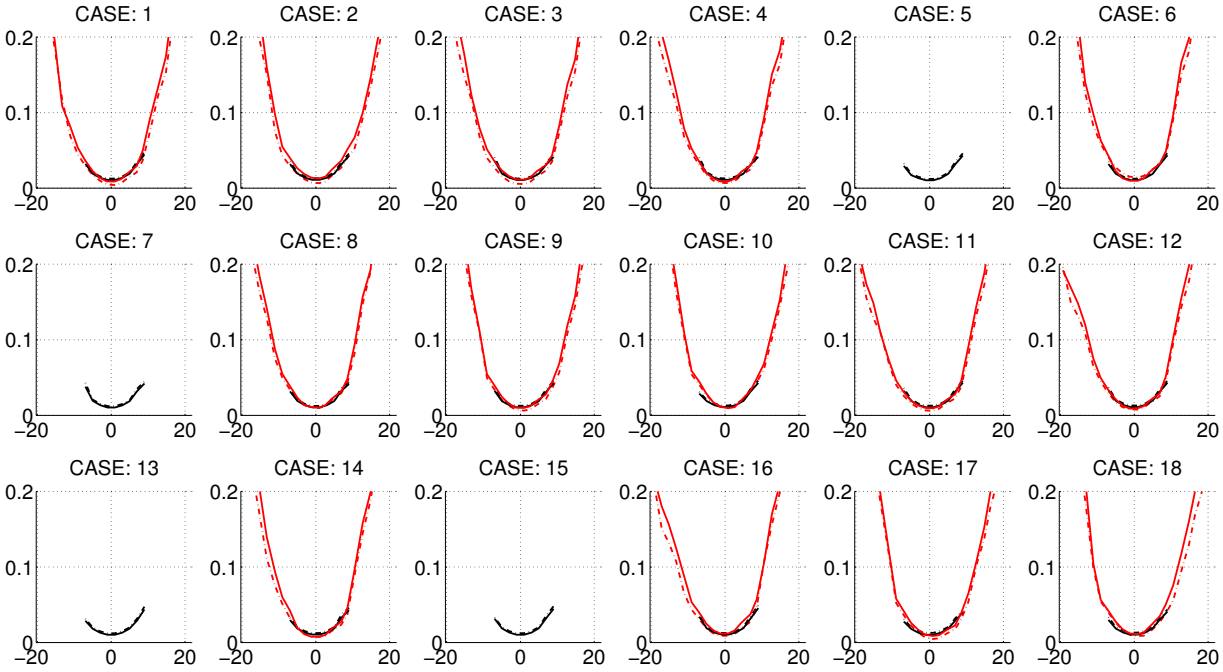

FIGURE 6.5: Drag coefficient versus α for both Reynolds numbers $4.5 \cdot 10^5$ (dot-dashed lines) and $5.5 \cdot 10^5$ (solid lines). Red lines represent the experimental database whereas black lines represent numerical database. Sample with only 18 variants detailed in Table B.1, Appendix B, of box-wing configuration out of the total 72.

in C_D are: MaxE=$7.3 \cdot 10^{-1}$ and RMSE=$1.2 \cdot 10^{-1}$. Figure 6.6 shows these differences evaluated along the range of angles of attack, instead. As expected, near the zero angle of attack, errors are smaller. It needs to be emphasized that both methods are tools used to analyse aerodynamics in very diverse phases of the design process. The numerical method is considered as time-efficient and robust but fails to be accurate and reliable, aspects strongly required for advanced design phases where wind tunnel testing is needed. Therefore, the observed differences can be assumed and the numerical database can be used as a backup check when the experimental missing data are reconstructed.

6.4 Gappy reconstruction in the experimental database

The scope of this section is studying and evaluating the quality of the gappy reconstruction for the experimental database using the gappy-noisy-HOSVD method. In the following, errors are normalized in the same way as in §4 and §5 to facilitate their comparison.

As mentioned, due to the lack of a complete experimental database, the numerical database is used instead as a check, even though its dimension referred to

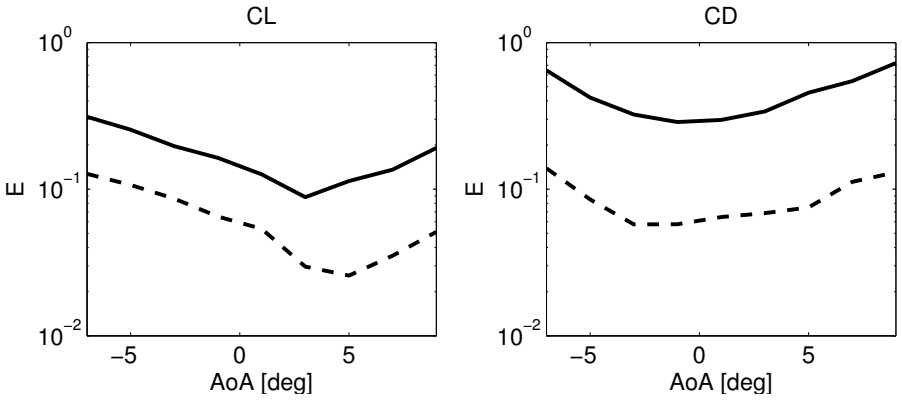

FIGURE 6.6: Lift (left plot) and drag (right plot) coefficient errors, plotting MaxE (solid lines) and RMSE (dashed lines), between the experimental and the numerical databases evaluated for each angle of attack taking as reference the numerical database.

as the angle of attack is smaller. This fact forces to work with a reduced experimental tensor, in terms of angle of attack, in order to be consistent in the analysis when both databases are compared.

The information in the experimental tensor is highly sparse since its gappyness reach 62.5% of the total tensor. This fact clearly makes difficult the reconstruction. In addition, experimental errors are assumed to be present and consequently, the corresponding database is assumed to be contaminated by some noise. The gappy-noisy-HOSVD method can be appropriately applied in these cases. Figures 6.7 and 6.8 show with blue lines the reconstructed missing lift and drag coefficients (respectively).

From the aerodynamic point of view, reconstructed data agree remarkably well with all the patterns observed in the non-gappy data, resulting from the numerical and experimental curves comparison. Some considerations can be highlighted:

- The baseline configuration uses a symmetric airfoil in all lifting surfaces, a characteristic that can be observed from the angle of attack where the flight condition of zero lift occurs. The Vortex Lattice solution exactly matches this expected behaviour. Notwithstanding that a small deviation is observed in the experimental database, the same shift can be systematically checked in all configurations which C_L values has been reconstructed.
- Wind tunnel results clearly exhibit a higher slope of $C_L - \alpha$ curves in comparison with the numerical solution. This pattern is also repeated in the reconstructed cases.
- Reynolds number influence on C_L is only noticeable in the case of the wind tunnel results. This fact is also evidenced by the same trends in the reconstructed cases.

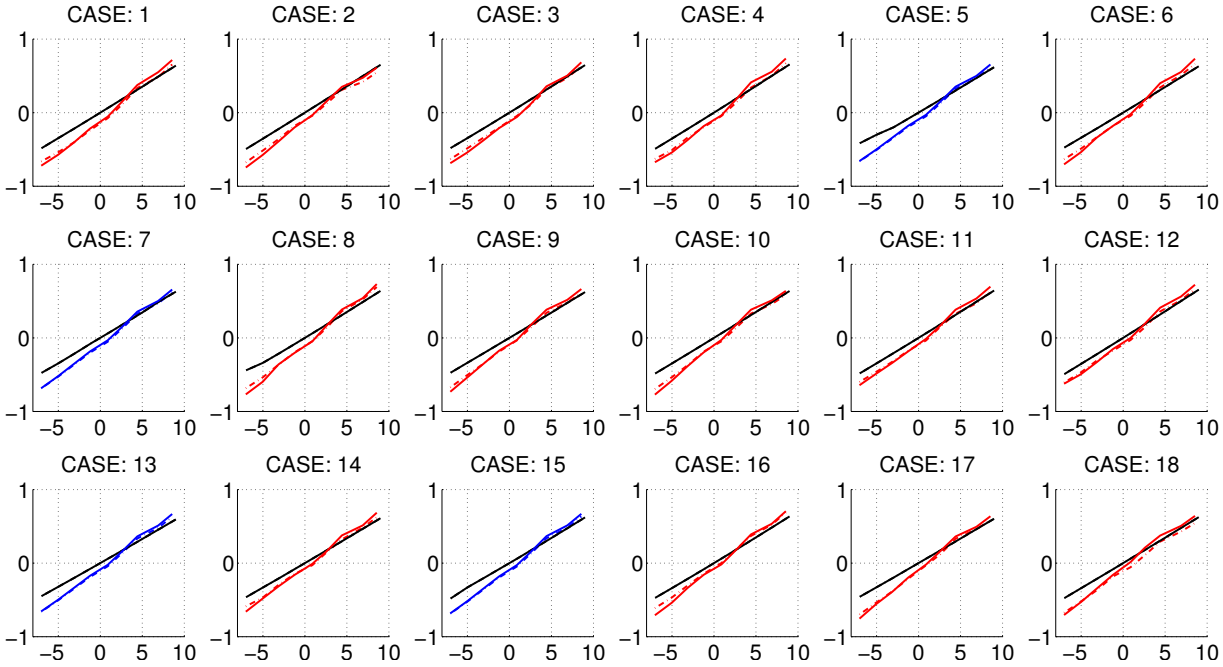

FIGURE 6.7: Lift coefficient curves versus α for both Reynolds numbers $4.5 \cdot 10^5$ (dot-dashed lines) and $5.5 \cdot 10^5$ (solid lines), when the gappy-noisy-HOSVD method has been applied. Red lines represent the experimental database in the non-gappy configurations, blue lines represent the reconstructed C_L values in the gappy configurations using 15 modes and black lines represent the numerical database. Sample with only 18 variants detailed in Table B.1, Appendix B, of box-wing configuration out of the total 72.

Drag coefficient curves comparison also demonstrates an excellent agreement between both databases. Some considerations can be highlighted:

- From Figure 6.8 it can be observed that both wind tunnel and numerical curves match, in almost all cases, very well in terms of C_{D0}. It should be recalled that the Vortex Lattice method does not consider viscous effects but its results has been corrected by an approximation predicted by a pseudo-3D boundary layer analysis. Regarding the reconstructed solution, the same behaviour is replicated. The matching for lift coefficient (Figure 6.7) was much worse as the numerical method does not include any effect due to thickness in the lift.
- As expected, experimental C_D values are slightly higher than numerical ones. Experimental measurement of drag is very sensitive in terms of surface quality of the model, flow quality in the test chamber, bumps and hollows for holding the model, etc. Although supports correction was applied during the wind tunnel data post processing, this relative higher value in drag is acceptable. The reconstruction seems consistent with this observation.

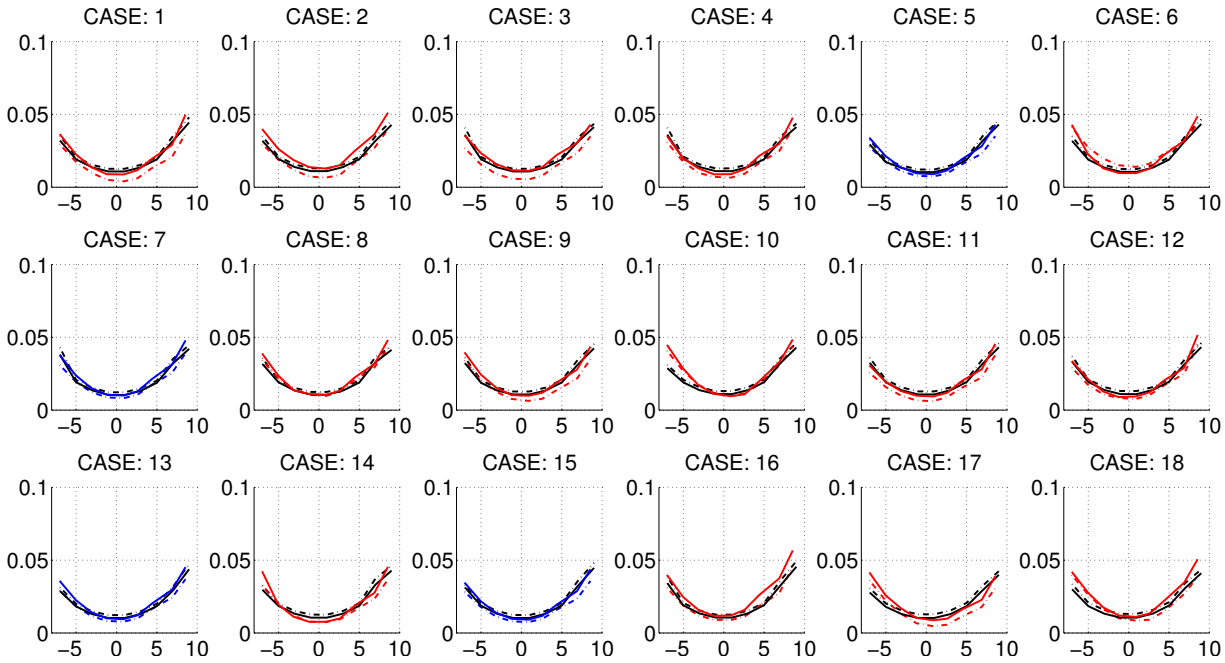

FIGURE 6.8: Drag coefficient curves versus α for both Reynolds numbers $4.5 \cdot 10^5$ (dot-dashed lines) and $5.5 \cdot 10^5$ (solid lines), when the gappy-noisy-HOSVD method has been applied. Red lines represent the experimental database in the non-gappy configurations, blue lines represent the reconstructed C_D values in the gappy configurations using 12 modes and black lines represent numerical database. Sample with only 18 variants detailed in Table B.1, Appendix B, of box-wing configuration out of the total 72.

- As mentioned, the aerodynamic behaviour at highest (lowest) angles of attack is differently predicted by each method. Flow separation and stall is detected and measured in the wind tunnel database while the numerical solution cannot predict it. This feature can be slightly observed in both gappy and non-gappy regions because they display a more convex parabola than the numerical one.
- Reynolds number influence is always appreciable in C_D curves although the analysed values are very close and tendencies seem to show irregularities in the tested cases. This feature does not allow qualitatively evaluate the reconstruction under this perspective.

From a quantitative analysis point of view, error analysis in Figure 6.9-left, corresponding with C_L database reconstruction, exhibits constant behaviour when increasing the selected number of modes. MaxE and RMSE are of the same order (MaxE$\sim 10^{-1}$, RMSE$\sim 10^{-2}$) than those obtained from comparing both experimental and numerical C_L values beforehand. It means that the gappy-noisy-HOSVD method has not increased the already existent base error between the two databases which is mainly due to their different nature. Both the gappy and non-gappy error curves overlap which confirm this conclusion.

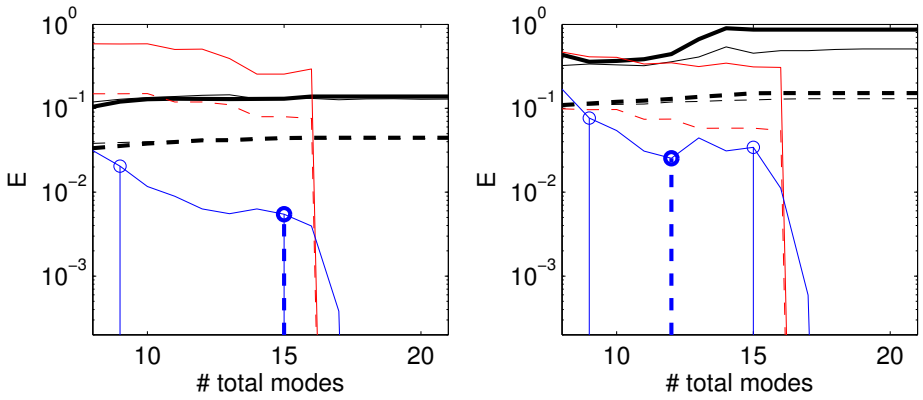

FIGURE 6.9: Counterpart of Figure **??** for both experimental C_L (left) and C_D (right) databases. Similarly, in these cases clean databases are unknown therefore the green lines (giving the errors comparing with the clean database) can not be plotted. Otherwise, black curves represent errors comparing with the numerical databases.

Figure 6.9-right shows MaxE and RMSE obtained for the C_D database reconstruction. Error curves are not monotone. Unlike left hand-side plot, the errors first decrease and then increase suggesting that drag coefficient database has some noise. Although both C_L and C_D databases have been extracted from same experiments, as mentioned before, drag measurement is quite sensitive to error influence in wind tunnel testing.

Notice that MaxE and RMSE exhibit again the same order of magnitude than the obtained from the preliminary comparison. It confirms that the method performance is fairly good considering how scatter the experimental tensors are. The quality of the reconstructed data are equated to the tested points.

6.5 Gappy reconstruction in the numerical database when same gappyness of the experimental database is present

For reference, same gappyness of the experimental databases has been applied to the numerical databases in order to exactly evaluate the performance of the HOSVD-based method with databases where missing data values are known.

These numerical databases can be considered as exact since they are free of noise. The reconstruction errors plots will be somewhat similar to the corresponding ones when 'clean' databases have been previously analysed. For this reason, the gappy-HOSVD method is applied instead in this case. Obviously, this change does neither affect the performance of the reconstruction nor its comparison with the subsection §6.4.

Counterparts of Figures 6.7 and 6.8 can be seen in Figures 6.10 and 6.11, respectively. The aerodynamic behaviour of missing cases are fairly well reconstructed in comparison with the original data (black lines) in both C_L and C_D databases (lines are almost undistinguishable). However, both MaxE and RMSE errors in the gappy regions, shown in Figure 6.12, exhibit a constant behaviour as the number of retained modes increases and of the same order than the obtained for the experimental databases reconstruction (see Figure 6.9). Even though the reconstruction is good and almost plot-indistinguishable from the original database, their high MaxE and RMSE errors are due to the excessive gappyness of the database (recall that 62.5% of the database is missing and it has a black-box nature for two of the 7 dimensions) and therefore, the method is not able to substantially reduce error values in this region. This observation is also confirmed by the MaxE and RMSE error curves obtained at the non-gappy region which considerably decrease as the number of retained modes increases.

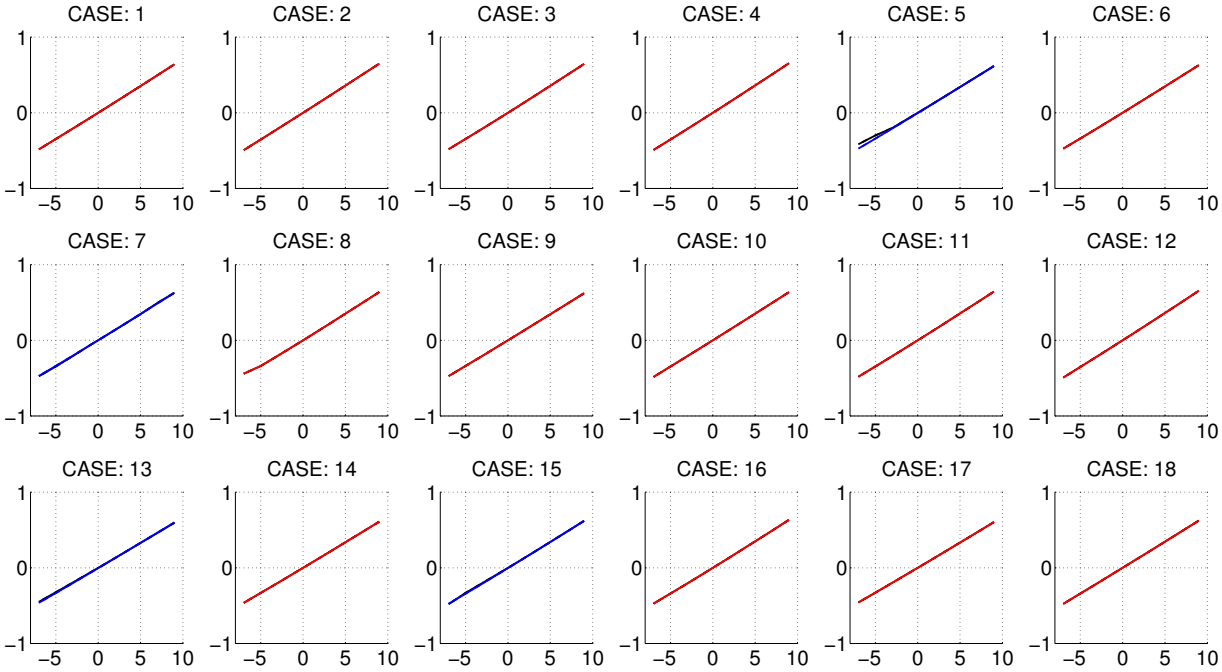

FIGURE 6.10: Counterpart of Figure 6.7 for the numerical database when same gappyness of the experimental database is present. The gappy region is reconstructed using 15 modes.

Subsequently, it seems reasonable to check the reconstruction quality when gradually decreases the number of gappy points. This study is intended to determine the influence the original gappyness type has on the performance of the method. As mentioned, the selection criterion of the tested configuration from the total map of 72 was not designed in order to obtain an optimal reconstruction of the missing data. Therefore, it is expected that increasing the number of tested configurations (reducing gappyness) the reconstruction errors significantly diminish, as Figure 6.13 confirms.

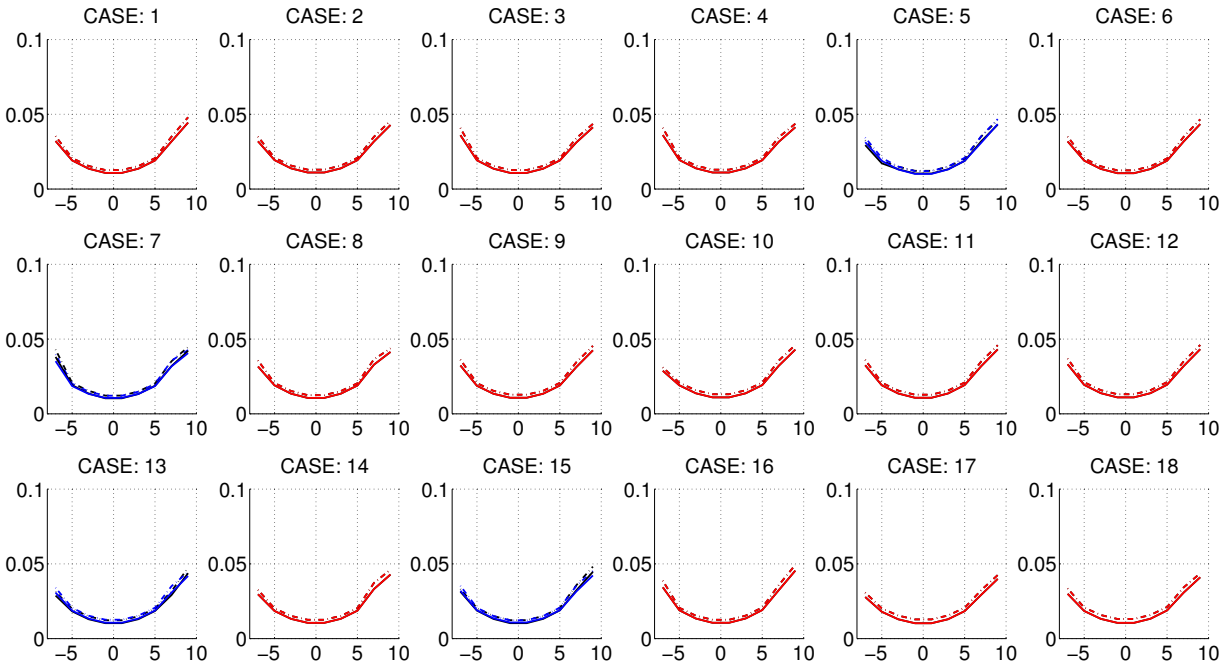

FIGURE 6.11: Counterpart of Figure 6.8 for the numerical database when same gappyness of the experimental database is present. The gappy region is reconstructed using 15 modes.

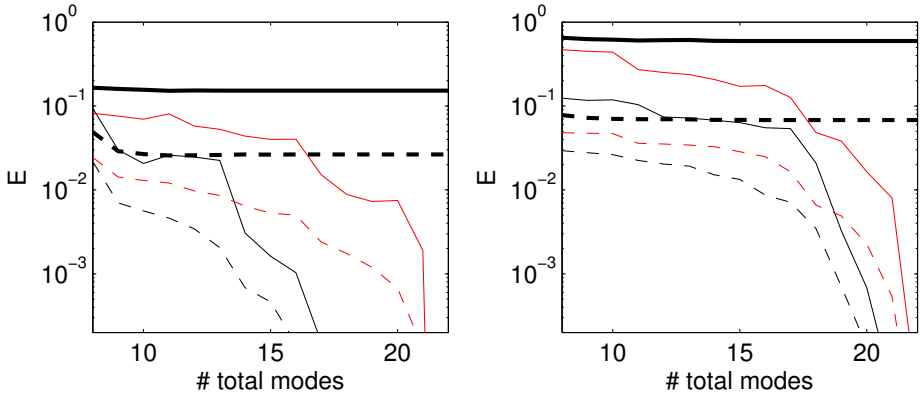

FIGURE 6.12: Counterpart of Figure 6.9 for the numerical database when same gappyness of the experimental database is present.

Last plot in Figure 6.13 corresponds to a database where only 7 combinations of the baseline configuration are missing, representing ∼10% gappyness. In general, errors have decreased considerably, as MaxE$= 3.4 \cdot 10^{-2}$ and RMSE$= 1.3 \cdot 10^{-2}$ in the gappy region using 13 modes, although the added configurations have been randomly selected. Obviously, an appropriate and systematic selection of the configurations to be tested could strongly improve the performance of the reconstruction.

Bringing up the last observation, a possible variant of the methods presented in this Thesis could come out to be able to choose which suitable configurations that have to be tested for the best reconstruction performance, checking all the

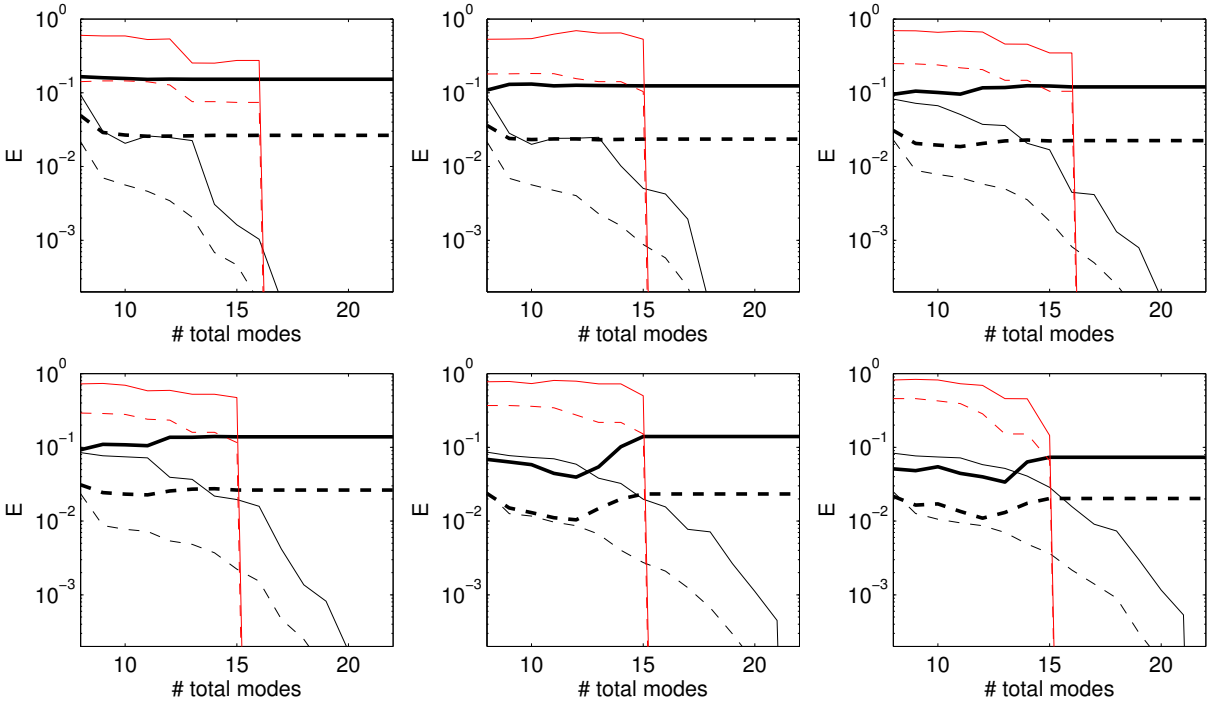

FIGURE 6.13: Counterpart of Figure 6.12 when the gappyness is reduced gradually. From the left to the right, downwards: the original experimental gappyness is present with 27 tested configurations (top-left, as a reference) and the followings refer to 35, 42, 50, 57 and 65 tested configurations, respectively (see Table B.1 in Appendix B).

configurations map.

6.6 Assessment of reconstruction performance using single-skipped-measured-data

A slightly different analysis, based using one of the measured points as a reference, for the reconstruction accuracy is detailed. As the experimental database is very sparse, a high number of tested configurations can not be skipped in order to be reconstructed with the proposed method and studying its performance. Instead, one tested data point is going to be omitted from the database and considered inside the gappy region to be reconstructed. This study allows to better analyse errors when the reconstructed data is compared with real experimental values instead of with a numerical approximation. Results obtained will be affected by same errors level as previous sections because the gappyness level in the experimental database is clearly high.

This procedure has been applied to all 27 tested configurations, one by one, in order to present a global analysis independent from what is the position of the omitted information inside the tensor. Error curves obtained for all 27 analysis

Region	MaxE	RMSE
Skipped config.	$1.30 \cdot 10^{-1}$ [0.7%]	$8.04 \cdot 10^{-2}$ [0.3%]
Skipped config. applying truncated HOSVD	$2.36 \cdot 10^{-1}$ [4.9%]	$1.46 \cdot 10^{-1}$ [2.7%]
Non-gappy region	$1.50 \cdot 10^{-1}$ [0.1%]	$7.94 \cdot 10^{-2}$ [0.0%]

TABLE 6.2: Gappy-noisy-HOSVD method performance when it is applied to the experimental database C_L and one tested configuration is omitted in addition to the inherent gappyness. MaxE and RMSE errors are presented in terms of mean values and variance (as percentage of the mean) obtained when same procedure is applied to 27 configurations.

Region	MaxE	RMSE
Skipped config.	$5.43 \cdot 10^{-1}$ [2.2%]	$2.68 \cdot 10^{-1}$ [0.8%]
Skipped config. applying truncated HOSVD	$2.57 \cdot 10^{-1}$ [7.5%]	$1.22 \cdot 10^{-1}$ [3.3%]
Non-gappy region	$7.09 \cdot 10^{-1}$ [0.1%]	$2.75 \cdot 10^{-1}$ [0.0%]

TABLE 6.3: Counterpart of Table 6.2 for the C_D database.

are similar to Figure 6.9, and are therefore omitted. Instead, Tables 6.2 and 6.3 present the MaxE and RMSE of the reconstructed configuration according to the estimate MSE, in terms of mean and variance values for all 27 tested configurations. Same regions where errors are usually measured in the database throughout this Thesis are detailed: the skipped configuration corresponds to the gappy region in these studied cases and consequently, the non-gappy region the rest of 26 configurations not skipped. For reference, the result of applying trucated HOSVD (retaining the same number of modes) to the original database is also included.

Reconstruction errors in Tables 6.2 and 6.3 exhibit a good performance of the gappy-noisy-HOSVD method considering that in these cases the gappyness level has increased until almost 64%. Although the actual error in the measurement is unknown, the method inherently filters that error out. Errors at the omitted configuration are of the same order than the non-gappy region errors, meaning that the reconstruction of the omitted configuration is the best we can get with such little information available.

Chapter 7

Conclusions and Future Work

A method has been developed to reconstruct industrial databases using a multidimensional extension of a seminal gappy reconstruction method introduced by Everson and Sirovich. The main idea has been to just replace plain POD (in fact, SVD) by HOSVD. On the other hand, in addition to some improvements already introduced for two-dimensional databases [33, 34], and other minor improvements to increase robustness, the method has been adapted to treat:

- Noisy databases, which is a major issue keeping realistic industrial databases in mind.
- Structured but non-rectangular databases, which are sometimes needed in specific applications, especially in Aerodynamics.

The methods developed in this Thesis have been applied, first to a toy model database (which has been used to both illustrate the application of the method and to calibrate it) and then to four aerodynamic databases, one resulting from a theoretical description, which could be considered as exact, another two obtained via CFD and the last one including experimental data from a wind tunnel test. The main conclusions are:

- The methods work very well when complexity is spread along the database, as in the toy model database. Aerodynamic databases show concentrated complexity and have demonstrated to be more demanding than other databases although the reconstructions are still robust and fairly good, especially when discretization errors are kept under control.
- The gappy-HOSVD method works very well when randomly distributed gappyness is present in the database. For concentrated gappyness, the performance of the methods is generally worse as the information is distributed less evenly. Furthermore, if the missing information is at the borders of the underlying physical domain, the efficiency of the method decreases as it needs to extrapolate. Nevertheless, MaxE and RMSE are about 6% and 0.9% respectively when reconstructing black-box gappyness at the boarder of the coarse toy model database, which is the worst case.
- Similarly, the methods give worse results when gappyness is concentrated in regions showing steep gradients.

- The gappy-HOSVD method can be applied to non-rectangular databases by embedding them in a larger rectangular database and treating the missing points as gappy points to reconstruct them. This allows to increase the accuracy in the interior region reconstruction in comparison to fixing the missing values to a constant number.
- The above statement opens the way for new strategies of gappyness treatment since the gappy-HOSVD method does not impose restrictions in the database form. The gappy-HOSVD method has been satisfactorily applied to portions of the complete database in order to reconstruct local gappyness even if by so doing some redundancies are lost in comparison to dealing with the entire database.
- When the database has intrinsic noise, a variant of the gappy-HOSVD method called gappy-noisy-HOSVD has been used to reconstruct the database filtering the noise interference. This is done by choosing the optimal number of retained modes. The method works well for the toy model and aerodynamic databases and can also clean the database from the noise. In general, the method gives comparable error levels for both types of gappyness distribution (random or concentrated) which suggests that the reconstruction is dominated by the presence of noise or similar artifacts which behave as noise.
- The methods are expected to perform better as the number of dimensions increases since redundancies in all directions are proliferated.

On the other hand, the results in This thesis suggest the following extensions and further applications of the method, which will be considered elsewhere:

- The method needs improvements when steep gradients (such as the suction jump near the leading edge) are present. An obvious way to solve the difficulty is to use a better mesh that concentrates points in the problematic regions. This has already been done to some extend above, but increasing further the mesh density near the leading edge would increase the CFD computational cost, which is not acceptable keeping in mind industrial applications. A possible computationally inexpensive solution would consist in combining the HOSVD method with a multiscale description, which is ahead the scope of this Thesis.
- Transonic and supersonic aerodynamics exhibit shock waves that are quite demanding for SVD-like methods (not to mention the more elaborate gappy-HOSVD methods developed in this Thesis). As further explained in [24] and [14], moving steep structures (in the considered ranges of the angle of attack, shock waves may move up to one third of the chord [14]) are not well described by SVD-like methods. This is because, in the end, the HOSVD reconstruction is a linear combination of the given data and a linear combination of moving jumps is not a jump, but a stair-like structure. In other words, moving jumps are even much more demanding than the static jumps considered in the last item. This difficulty may be

solved by using a specific treatment [24, 14], which is quite involved. The extension of these (or related) ideas in the context of the gappy-HOSVD methods is well ahead the scope of this Thesis.

- Even though the methods developed in this Thesis have only been tested in aerodynamic databases, they rely on general ideas and principles and thus they are expected to apply to more general multidimensional databases with a technological origin. The only condition is that appropriate redundancies be present along the various database dimensions.
- The proposed methods include tunable parameters to be setted. This Thesis has demonstrated that their performance is fairly insensitive to the values of these parameters since the same set of parameter values is used for all applications. In particular, the gappy-HOSVD method parameters have been studied in detail concluding that their incidence in the required computational cost is excessive with little effect in the accuracy of the method. On the other hand, the gappy-noisy-HOSVD method includes two new parameters which detailed study is proposed as future work. Although the method has been demonstrated to be robust, a general criterion how to select those values is an open problem.
- Applying the gappy-HOSVD method to a portion of the database instead of dealing with the entire one has been demonstrated to be useful but its efficiency is closely related to the division strategy. An in-depth study on different strategies of division and the overlapping amongst the portions is also proposed as future work. The results obtained in this Thesis preliminarily show that the proposed methods can be applied to huge databases keeping a reasonable computational efficiency.

Two further extensions of the methods are proposed:

- The gappy-noisy-HOSVD method is able to reconstruct multidimensional databases when the missing data are at known positions. Its behaviour suggests that, after appropriate modifications, it could be used to identify errors at unknown positions and then correct them by treating them as gappy points.
- Since, in general, the gappy-HOSVD method identifies patterns in the databases, the method can be adapted to identify critical points for a good quality database reconstruction. This can be used as a sampling method to, e.g., identify good sensor positioning in experimental tests.

Developing first idea, the gappy-noisy-HOSVD method can be further modified in order to treat the problem of error filtering, namely cleaning the database when this is contaminated with unknown errors at unknown positions. The basic idea is that, at each step in the iteration process, those database elements that show the largest deviation from the SVD (or HOSVD in higher order tensors) reconstruction are treated as gappy elements. The resulting method filters out well both $O(1)$ and small errors, both when these are random and when errors are systematic. This new method [42] is firstly explained in detail for

two dimensional databases, namely *cleaning-SVD method*, and then extended to multidimensional databases by substituting SVD by HOSVD. Several applications are included where the method has been successfully tested. In particular, aerodynamic databases generated by CFD analysis with random and systematic errors located both at randomly distributed and concentrated positions are treated.

According to the second idea and left as the object of future research, a further modification of the gappy-HOSVD method could be able to guess which points are critical in a database with the scope of using them to reproduce the rest of the values with the minimum number of tested points. This interesting application, based on the same principal as the filtering method, iteratively analyse elements that show highest errors between two consecutive HOSVD reconstructions and thus are main candidates to be selected.

Appendix A

Theoretical pressure coefficient distribution on an airfoil

The calculation of the pressure distribution over a wing section in inviscid flow is a classical problem of aerodynamics. Several methods have been developed for characterizing this two-dimensional problem. A well known semi-empirical method by Theodorsen (compiled in Abbott's work [55]) is used, based on a conformal mapping that transforms an arbitrary airfoil surface onto the surface of a circular cylinder. The flow field around a circular cylinder with circulation in presence of an uniform stream of ideal fluid is a basic flow pattern widely characterized. Mainly, the local velocities over this surface can be calculated and then, the pressure is obtained by means of Bernoulli's equation.

A conformal mapping consists on a transformation function that preserves angles locally between the domain and its image in such a manner that the detailed shape of infinitesimal elements of area is not changed, but not necessarily their size or curvature. This tool is widely used in bidimensional aerodynamics for studying the flow past an airfoil. One of such conformal mapping is the Joukowsky mapping: an airfoil from the Joukowsky family given in a complex ς plane is mapped to a circle in the complex z' plane by the inverse transformation in Eq. (A.1). But if this transformation is applied to an arbitrary airfoil, the resulting curve in the z' plane will therefore have a nearly circular shape.

$$f : \varsigma \to z' \quad \text{with} \quad \varsigma, z' \subseteq \mathbb{C}^2 \qquad \varsigma = z' + \frac{1}{z'} \tag{A.1}$$

Theodorsen, based on a Joukowsky-like transformation, devised a two-step transformation, represented in Figure A.1, which maps the nearly circular curve in the z' (referred to as near circle), obtained by Eq. (A.1), into a true circle. Thus, the flow around the near circle and hence around the airfoil, can be derived from the flow around the true circle by a rapidly converging process divided in three steps:

1. Derivation of relations between the flow field of the plane of the airfoil (ς plane) and the near circle plane (z' plane).
2. Derivation of relations between the flow field of the z' plane and the true circle plane (z plane).

3. Combination of the above relations in order to obtain the final expression for the velocity distribution in the ς plane in terms of the coordinates (x,y) of the airfoil.

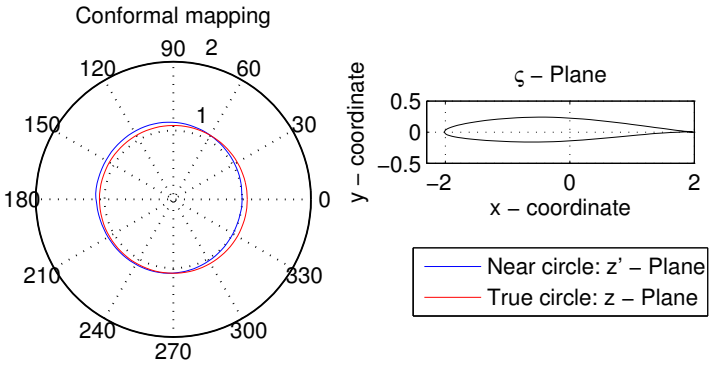

FIGURE A.1: Representation of the two derived planes z and z' used for mapping the NACA 64-210 airfoil from ς plane to a circle in the z plane.

As mentioned, the conformal mapping between the z' and the ς plane in Eq. (A.1) allows to obtain the relations between coordinates of the corresponding planes (see Eq. (A.2)). The coordinates of the z' plane, ψ and θ ($z' = e^{\psi+i\theta}$) can be expressed in terms of x and y (coordinates of $\varsigma = x + iy$), as well as the factor that relates velocities in the z' plane to those in the ς plane, ${}^{dz'}/_{d\varsigma}$. The coordinate ψ can be written as a function of θ, $\psi(\theta)$ where $\theta(x, y)$. This step has been widely addressed in literature, thus details are omitted for the sake of brevity.

$$\begin{aligned} x &= 2cosh(\psi) \cdot cos(\theta) \\ y &= 2sinh(\psi) \cdot sin(\theta) \end{aligned} \tag{A.2}$$

The coordinates (x,y) of the airfoil are defined in such a way that the point midway between the nose of the wing section and its centre of curvature is $(-2, 0)$ and the trailing edge (in the case of having sharp trailing edges) is $(2, 0)$, as can be seen in Figure A.1-right. The x-axis corresponds with the line joining these two points. The y-axis is defined as a Cartesian coordinate system.

According to the second step, the conformal mapping between the z' plane and the z plane is

$$z' = z \cdot e^{\sum_{n=1}^{\infty}(A_n+iB_n)({}^{1}/_{z^n})} \tag{A.3}$$

where the coordinates of the z plane are defined as $z = e^{\lambda+i\varphi}$. The mapping relations between both pair of coordinates can be derived from Eq. (A.3), considering that $z' = z \cdot e^{\psi-\lambda+i(\theta-\varphi)}$. Defining z in a polar coordinate-like form $z = r \cdot (cos(\varphi) + isin(\varphi))$ in the obtained expression, $\psi - \lambda$ and $\theta - \varphi$ become

expressed by two Fourier expansions as conjugate functions:

$$\begin{cases} \psi - \lambda = \sum_{n=1}^{\infty} \left(\frac{A_n}{r^n} cos(n\varphi) + \frac{B_n}{r^n} sin(n\varphi) \right) \\ \theta - \varphi = \sum_{n=1}^{\infty} \left(\frac{B_n}{r^n} cos(n\varphi) - \frac{A_n}{r^n} sin(n\varphi) \right) \end{cases} \tag{A.4}$$

In the z' plane, the near circle is defined by a varying radius e^{ψ} which mean, namely e^{ψ_0}, is imposed to be the true circle radius e^{λ} in order to minimize the variance of $(\psi - \lambda)$.

$$\lambda = \psi_0 = \frac{1}{2\pi} \int_0^{2\pi} \psi d\theta \tag{A.5}$$

The relation $\psi(\theta)$ has been previously obtained, so ψ_0 can be easily calculated and hence λ in Eq. (A.4) is known. From first relation in Eq. (A.4), ψ becomes expressed as a function of φ, and the independent term is already known. Here, the function ψ is approximated by a finite trigonometric series of the form

$$\psi(\varphi) = \psi_0 + \sum_{m=1}^{n-1} (A_m cos(m\varphi) + B_m sin(m\varphi)) + A_n cos(n\varphi) \tag{A.6}$$

where, ψ is known as a discrete function defined at the points $\psi_r(r\pi/n)$ where $r = 0, ..., (2n-1)$, in the range $0 \leq \varphi \leq 2\pi$. Therefore, coefficients A_m, B_m and A_n in Eq. (A.6) can be calculated.

Now, with this in mind, the second relation in Eq. (A.4), $\theta - \varphi$, can be addressed. For convenience, a new variable ϵ is introduced such as $\epsilon(\varphi) = \varphi - \theta$. In the same manner as with ψ, $\epsilon(\varphi)$ can be expressed as the conjugate discrete function of Eq. (A.6) where the coefficients are already known.

$$\varphi - \theta = \epsilon(\varphi) = \sum_{m=1}^{n-1} (A_m sin(m\varphi) - B_m cos(m\varphi)) + A_n sin(n\varphi) \tag{A.7}$$

Substituting the known coefficients and evaluating $\epsilon(\varphi)$ at the same points $\varphi = r'\pi/n$ ($r' = 0, ..., (2n-1)$) where ψ is known, the expression in Eq. (A.7) can be rearranged as:

$$\epsilon(\varphi) = -\frac{1}{n} \sum_{K=1}^{2n-1} \left(\psi_K cot\left(\frac{K\pi}{2n} \right) \right) \quad \text{where} \quad \psi_K = \psi\left(\varphi + \frac{K\pi}{n} \right), \quad K \text{ is odd.} \tag{A.8}$$

It should be noticed that, in practice, ψ is known as a function of θ (obtained from step 1). As ϵ is discretized in the complete range $0 \leq \psi \leq 2\pi$, the expression $\epsilon(\theta)$ can be obtained straightforward.

Now, the relation between the flow field in the z' and z plane can be established. The factor relating the velocities between the near circle and the true circle plane $\frac{dz'}{dz}$ can be obtained by deriving the already known relation between both planes $z' = z \cdot e^{\psi - \psi_0 + i(\theta - \varphi)}$. Rearranging the derived expression and substituting $\epsilon =$

$\varphi - \theta$, it can be written as

$$\frac{dz'}{dz} = \frac{z'}{z}\frac{1 - i(d\psi/d\theta)}{1 + (d\epsilon/d\theta)} \tag{A.9}$$

From the already known expressions $\psi(\theta)$ and $\epsilon(\theta)$, $d\psi/d\theta$ and $d\epsilon/d\theta$ can be calculated. These derivatives should be numerically computed as terms of Eq. (A.9), where a central differencing scheme has been used.

Once the relationship between the airfoil plane and the true circle plane (steps 1 and 2) are found, the step 3 can be done. The flow past the true circle of radius e^{ψ_0}, in z plane, is described by the Eq. (A.10) that comes from the widely known theory about two-dimensional flow past a circular cylinder with circulation.

$$w = V\left(z + \frac{e^{2\psi_0}}{z}\right) + \frac{i\Gamma}{2\pi} ln \frac{z}{e^{\psi_0}} \tag{A.10}$$

The velocities in the true circle plane $^{dw}/_{dz}$ can be obtained from Eq. (A.10).

Rearranging the expressions from the conformal mapping, the potential theory solution can be obtained but it does not fulfil the Kutta-Joukowsky condition. Unreasonable velocities around the trailing edge should be avoided by forcing the rear stagnation point of the true circle plane to correspond to the trailing edge of the airfoil plane.

The flow past the circle is rotated through an angle α_0 equal to the angle of attack of the wing section. Due to circulation Γ, the rear stagnation point is shifted from $\theta = \pi$, and the corresponding point in the z plane to the trailing edge is situated at $z = e^{\psi_0 + i(\alpha_0 + \epsilon_T)}$. It is therefore necessary to rotate the stagnation point by an amount $\alpha_0 + \epsilon_T$ and consequently the circulation becomes defined in order to satisfy this condition.

Finally, last step can be concluded with obtaining the velocities in the airfoil plane from the already calculated velocities in the true circle plane, in Eq. (A.10) by inverting the conformal transformations:

$$\frac{dw}{d\varsigma} = \frac{dw}{dz}\frac{dz}{dz'}\frac{dz'}{d\varsigma} \tag{A.11}$$

After substituting corresponding factors in Eq. (A.11), the complete expression Eq. (A.12) is

$$\left|\frac{dw}{d\varsigma}\right| = v = V\frac{[1 + (^{d\epsilon}/_{d\theta})]e^{\psi_0}}{\sqrt{(sinh^2(\psi) + sin^2(\theta))[1 + (^{d\psi}/_{d\theta})^2]}}[sin(\theta + \alpha_0 + \epsilon) + sin(\alpha_0 + \epsilon_T)] \tag{A.12}$$

where v is the local velocity at any point on the airfoil surface and V is the free-stream velocity.

The pressure coefficient, Eq. (A.13), can be obtained by applying Bernoulli's equation throughout the flow field where v/V is known by the Eq. (A.12).

$$C_p = 1 - \frac{v^2}{V^2} \tag{A.13}$$

Eq. (A.13) represents the pressure coefficient evaluated at wing section surface

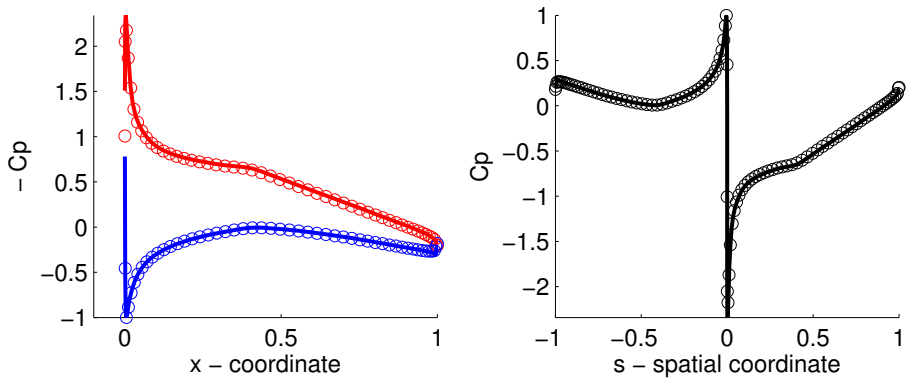

FIGURE A.2: Comparison between XFOIL inviscid solution (solid lines) and the calculated discretization of the theoretical C_p distribution (circle symbols) for the NACA 64-210 airfoil plotted in Cartesian coordinate system (left) and s-spatial coordinate (right). Left plot shows both the suction side (red) and the pressure side (blue).

points which Cartesian coordinates are known as (x, y), as shown in Figure A.2-left. Throughout this Thesis, the variable x will be transformed in s, called spatial coordinate, which represent the arch length along the cross section. It increases from the trailing edge at the pressure side, $-s_{te}$, to the trailing edge at the suction side, s_{te}. The pressure coefficient distribution over s-coordinate is shown in Figure A.2-right. Both figures also compare the obtained C_p distribution with the inviscid solution obtained by using XFOIL tool [56], with a very good agreement. The inviscid formulation of XFOIL is based on a simple linear-vorticity stream function panel method. The most important requirement for building the theoretical database is exactly evaluating the C_p values at certain points of the airfoil surface, where in some cases are not equispaced or very close between them along the s-coordinate. The selected method has been used instead of the XFOIL tool in order to avoid anomalous behaviour of this tool when panel sizes are not suitably distributed for an efficient performance of it.

Appendix B

Experimental and numerical databases reconstruction

The experimental database and its numerical calculated counterpart, treated in chapter §6, are composed by the aerodynamic coefficients C_L and C_D of 72 geometries tested at different Reynolds numbers and a wide range of angles of attack. Five geometric parameters and its combinations define the complete map of 72 studied geometries as variations of a baseline box-wing configuration, referred to as in Tables B.1 and B.2.

This appendix includes the results for each analysis of the mentioned chapter, for all 72 cases.

Figures B.1 and B.2 include the counterparts of Figure 6.4, and Figures B.3 and B.4 the counterparts of Figure 6.5, for all 72 geometries, corresponding with the experimental database analysis.

Figures B.5 and B.6 include the counterparts of Figure 6.7, and Figures B.7 and B.8 the counterparts of Figure 6.8, for all 72 geometries, corresponding with the gappy reconstruction analysis in the experimental database.

Figures B.9 and B.10 include the counterparts of Figure 6.10, and Figures B.11 and B.12 the counterparts of Figure 6.11, for all 72 geometries corresponding with the gappy reconstruction analysis in the numerical database when same gappyness of the experimental database is present.

	L/c	G/b	SW_{fr}	SW_{re}	Dih
CASE: 1	2	0.060	0°	0°	0.0°
CASE: 2	2	0.085	0°	0°	0.0°
CASE: 3	3	0.060	0°	0°	0.0°
CASE: 4	3	0.085	0°	0°	0.0°
CASE: 5	2	0.060	0°	0°	2.5°
CASE: 6	2	0.085	0°	0°	2.5°
CASE: 7	3	0.060	0°	0°	2.5°
CASE: 8	3	0.085	0°	0°	2.5°
CASE: 9	2	0.060	0°	23°	0.0°
CASE: 10	2	0.085	0°	23°	0.0°
CASE: 11	3	0.060	0°	23°	0.0°
CASE: 12	3	0.085	0°	23°	0.0°
CASE: 13	2	0.060	0°	23°	2.5°
CASE: 14	2	0.085	0°	23°	2.5°
CASE: 15	3	0.060	0°	23°	2.5°
CASE: 16	3	0.085	0°	23°	2.5°
CASE: 17	2	0.060	0°	40°	0.0°
CASE: 18	2	0.085	0°	40°	0.0°
CASE: 19	3	0.060	0°	40°	0.0°
CASE: 20	3	0.085	0°	40°	0.0°
CASE: 21	2	0.060	0°	40°	2.5°
CASE: 22	2	0.085	0°	40°	2.5°
CASE: 23	3	0.060	0°	40°	2.5°
CASE: 24	3	0.085	0°	40°	2.5°
CASE: 25	2	0.060	23°	0°	0.0°
CASE: 26	2	0.085	23°	0°	0.0°
CASE: 27	3	0.060	23°	0°	0.0°
CASE: 28	3	0.085	23°	0°	0.0°
CASE: 29	2	0.060	23°	0°	2.5°
CASE: 30	2	0.085	23°	0°	2.5°
CASE: 31	3	0.060	23°	0°	2.5°
CASE: 32	3	0.085	23°	0°	2.5°
CASE: 33	2	0.060	23°	23°	0.0°
CASE: 34	2	0.085	23°	23°	0.0°
CASE: 35	3	0.060	23°	23°	0.0°
CASE: 36	3	0.085	23°	23°	0.0°

TABLE B.1: List of numered cases representing half map of geometries: from case #1 to #36. All tested at different Reynolds numbers and a wide range of angles of attack. The geometric parameter ranges are detailed in the Table 6.1.

	L/c	G/b	SW_{fr}	SW_{re}	Dih
CASE: 37	2	0.060	23°	23°	2.5°
CASE: 38	2	0.085	23°	23°	2.5°
CASE: 39	3	0.060	23°	23°	2.5°
CASE: 40	3	0.085	23°	23°	2.5°
CASE: 41	2	0.060	23°	40°	0.0°
CASE: 42	2	0.085	23°	40°	0.0°
CASE: 43	3	0.060	23°	40°	0.0°
CASE: 44	3	0.085	23°	40°	0.0°
CASE: 45	2	0.060	23°	40°	2.5°
CASE: 46	2	0.085	23°	40°	2.5°
CASE: 47	3	0.060	23°	40°	2.5°
CASE: 48	3	0.085	23°	40°	2.5°
CASE: 49	2	0.060	40°	0°	0.0°
CASE: 50	2	0.085	40°	0°	0.0°
CASE: 51	3	0.060	40°	0°	0.0°
CASE: 52	3	0.085	40°	0°	0.0°
CASE: 53	2	0.060	40°	0°	2.5°
CASE: 54	2	0.085	40°	0°	2.5°
CASE: 55	3	0.060	40°	0°	2.5°
CASE: 56	3	0.085	40°	0°	2.5°
CASE: 57	2	0.060	40°	23°	0.0°
CASE: 58	2	0.085	40°	23°	0.0°
CASE: 59	3	0.060	40°	23°	0.0°
CASE: 60	3	0.085	40°	23°	0.0°
CASE: 61	2	0.060	40°	23°	2.5°
CASE: 62	2	0.085	40°	23°	2.5°
CASE: 63	3	0.060	40°	23°	2.5°
CASE: 64	3	0.085	40°	23°	2.5°
CASE: 65	2	0.060	40°	40°	0.0°
CASE: 66	2	0.085	40°	40°	0.0°
CASE: 67	3	0.060	40°	40°	0.0°
CASE: 68	3	0.085	40°	40°	0.0°
CASE: 69	2	0.060	40°	40°	2.5°
CASE: 70	2	0.085	40°	40°	2.5°
CASE: 71	3	0.060	40°	40°	2.5°
CASE: 72	3	0.085	40°	40°	2.5°

TABLE B.2: List of numered cases representing half map of geometries: from case #37 to #72. All tested at different Reynolds numbers and a wide range of angles of attack. The geometric parameter ranges are detailed in the Table 6.1.

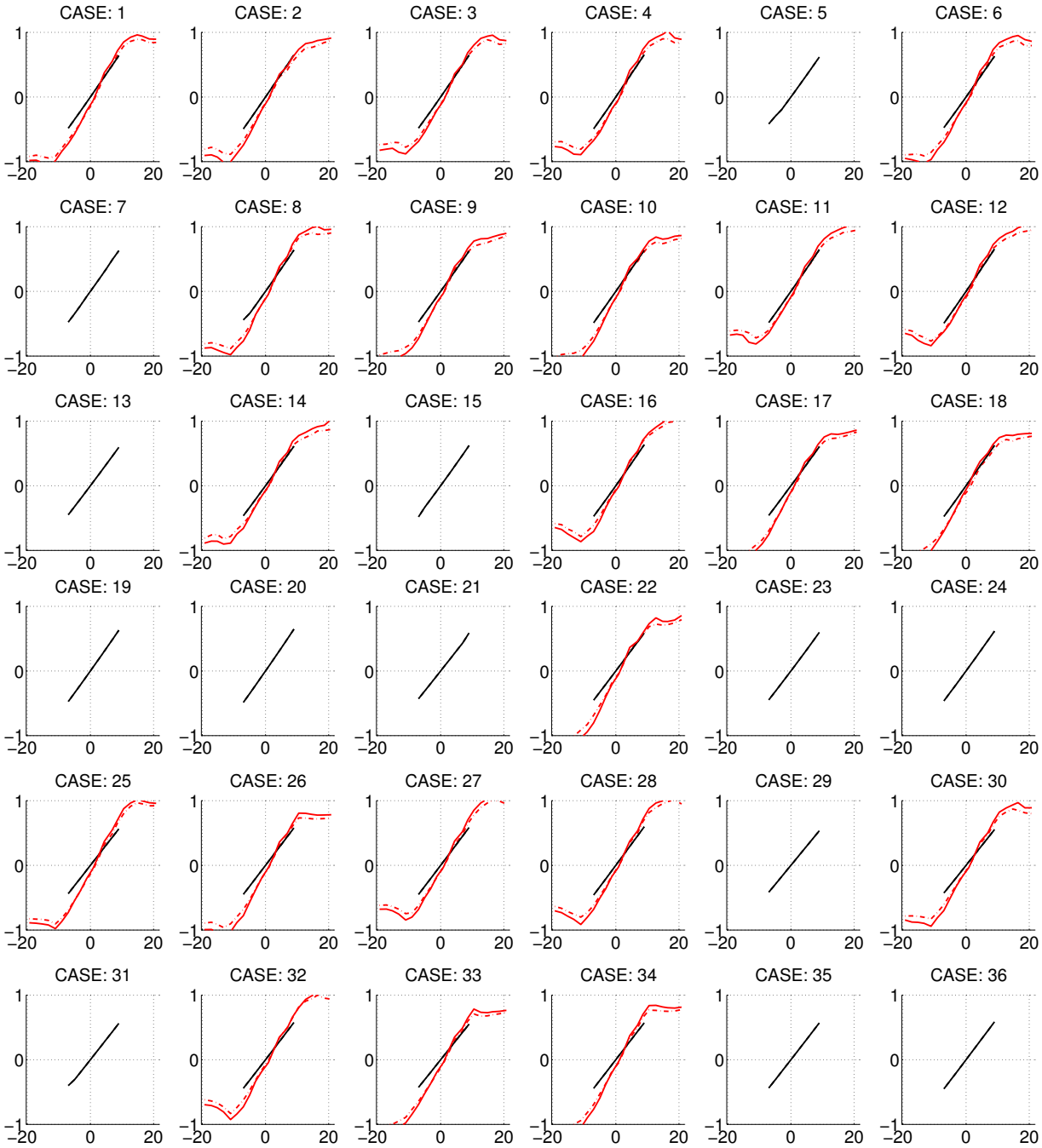

FIGURE B.1: Lift coefficient curves versus α for both Reynolds numbers $4.5 \cdot 10^5$ (dot-dashed lines) and $5.5 \cdot 10^5$ (solid lines). Red lines represent the experimental database whereas black lines represent numerical database (Reynolds number independent). Half map of geometries: from case #1 to #36.

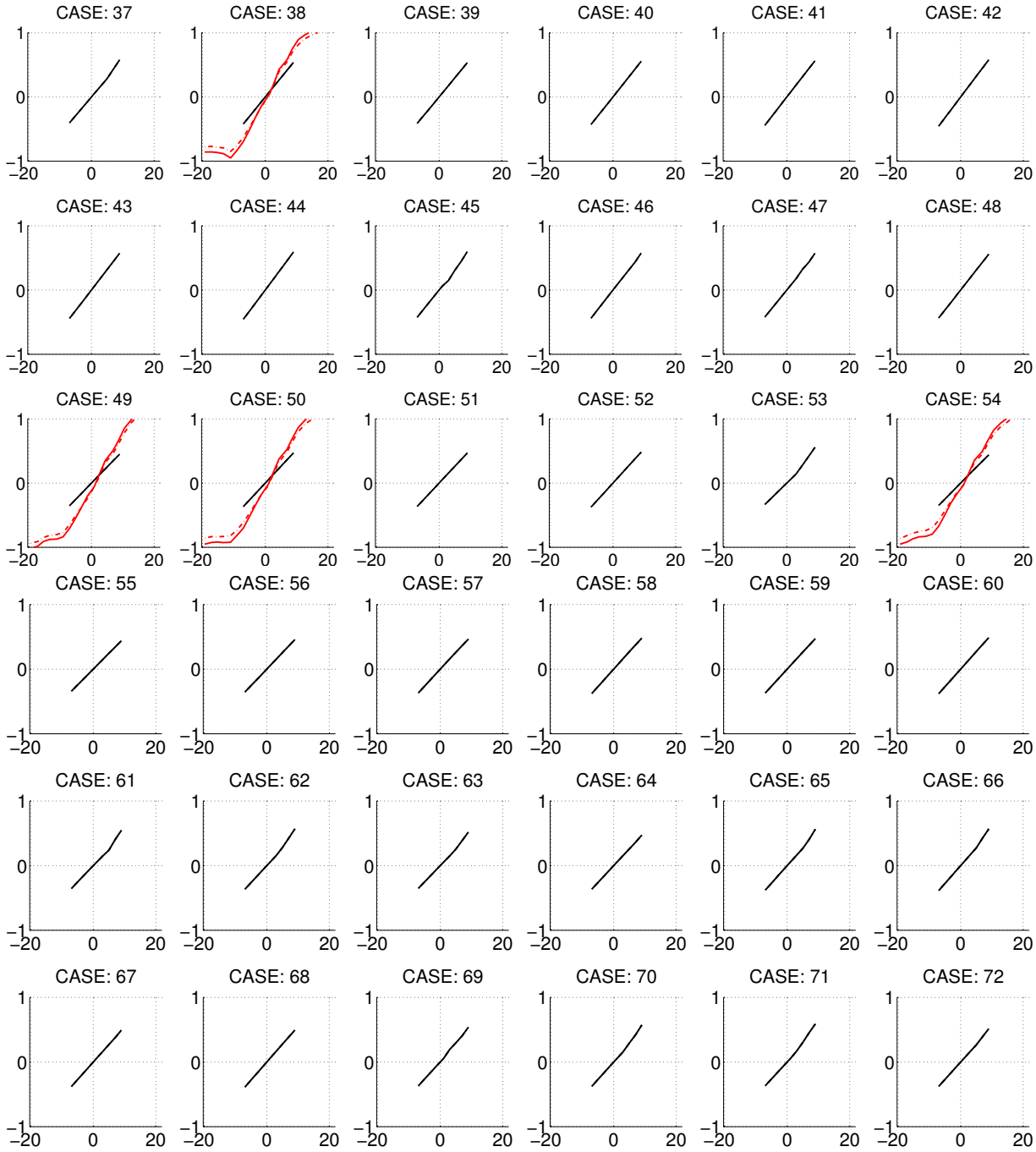

FIGURE B.2: Lift coefficient curves versus α for both Reynolds numbers $4.5 \cdot 10^5$ (dot-dashed lines) and $5.5 \cdot 10^5$ (solid lines). Red lines represent the experimental database whereas black lines represent numerical database (Reynolds number independent). Half map of geometries: from case #37 to #72.

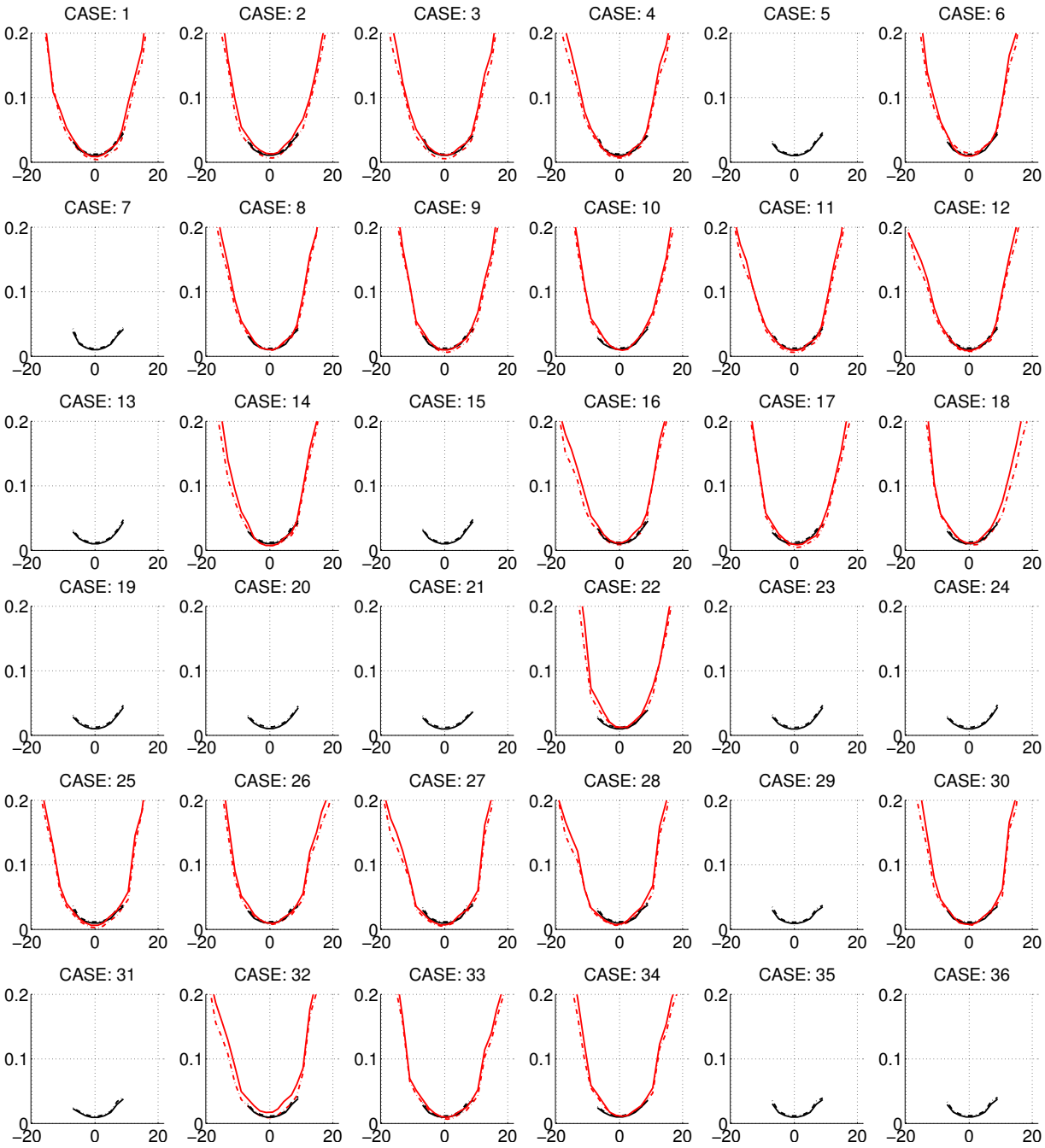

FIGURE B.3: Drag coefficient curves versus α for both Reynolds numbers $4.5 \cdot 10^5$ (dot-dashed lines) and $5.5 \cdot 10^5$ (solid lines). Red lines represent the experimental database whereas black lines represent numerical database. Half map of geometries: from case #1 to #36.

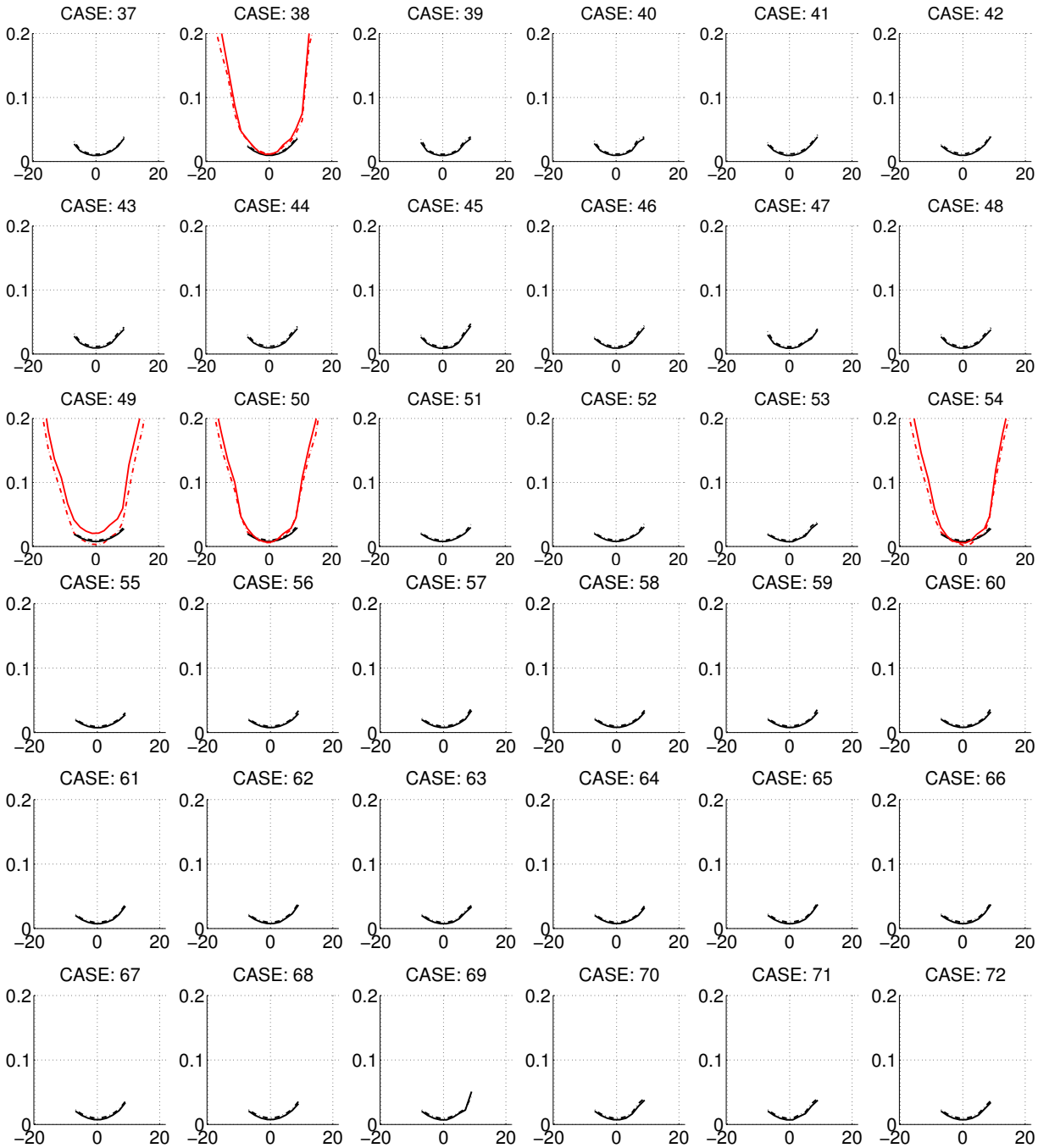

FIGURE B.4: Drag coefficient curves versus α for both Reynolds numbers $4.5 \cdot 10^5$ (dot-dashed lines) and $5.5 \cdot 10^5$ (solid lines). Red lines represent the experimental database whereas black lines represent numerical database. Half map of geometries: from case #37 to #72.

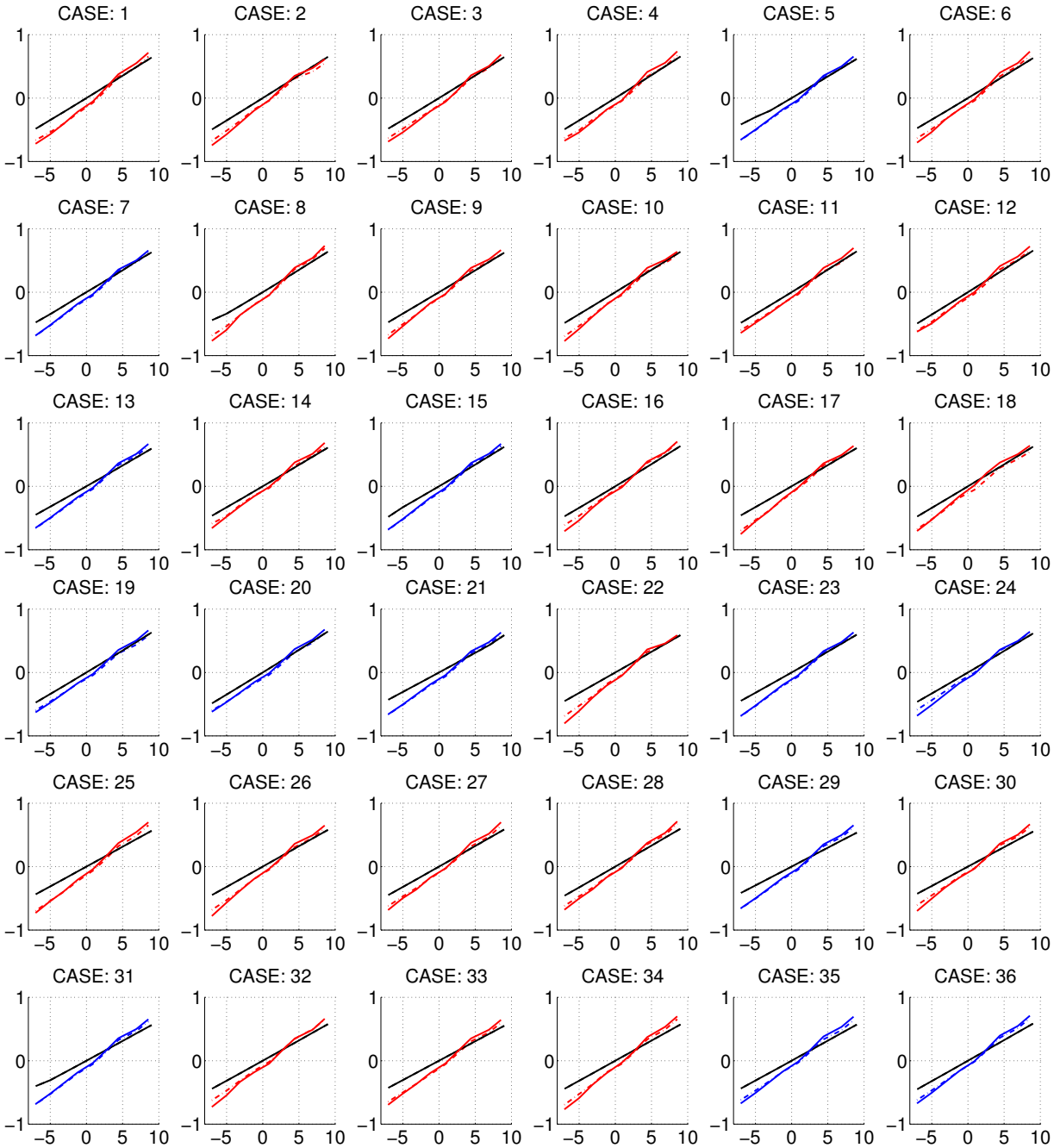

FIGURE B.5: Lift coefficient curves versus α for both Reynolds numbers $4.5 \cdot 10^5$ (dot-dashed lines) and $5.5 \cdot 10^5$ (solid lines), when the gappy-noisy-HOSVD method has been applied. Red lines represent the experimental database in the non-gappy configurations, blue lines represent the reconstructed C_L values in the gappy configurations using 15 modes and black lines represent numerical database. Half map of geometries: from case #1 to #36.

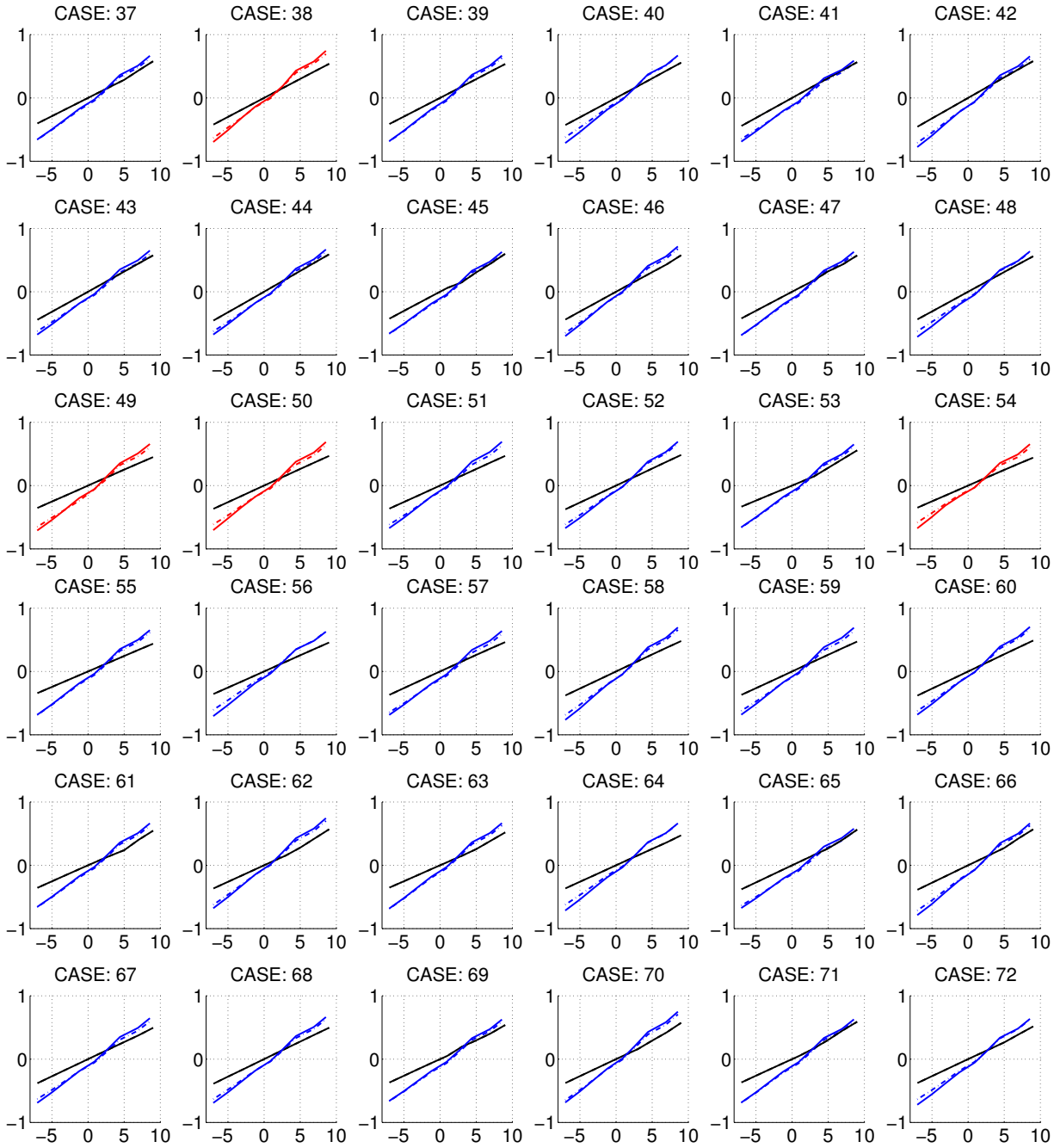

FIGURE B.6: Lift coefficient curves versus α for both Reynolds numbers $4.5 \cdot 10^5$ (dot-dashed lines) and $5.5 \cdot 10^5$ (solid lines), when the gappy-noisy-HOSVD method has been applied. Red lines represent the experimental database in the non-gappy configurations, blue lines represent the reconstructed C_L values in the gappy configurations using 15 modes and black lines represent numerical database. Half map of geometries: from case #37 to #72.

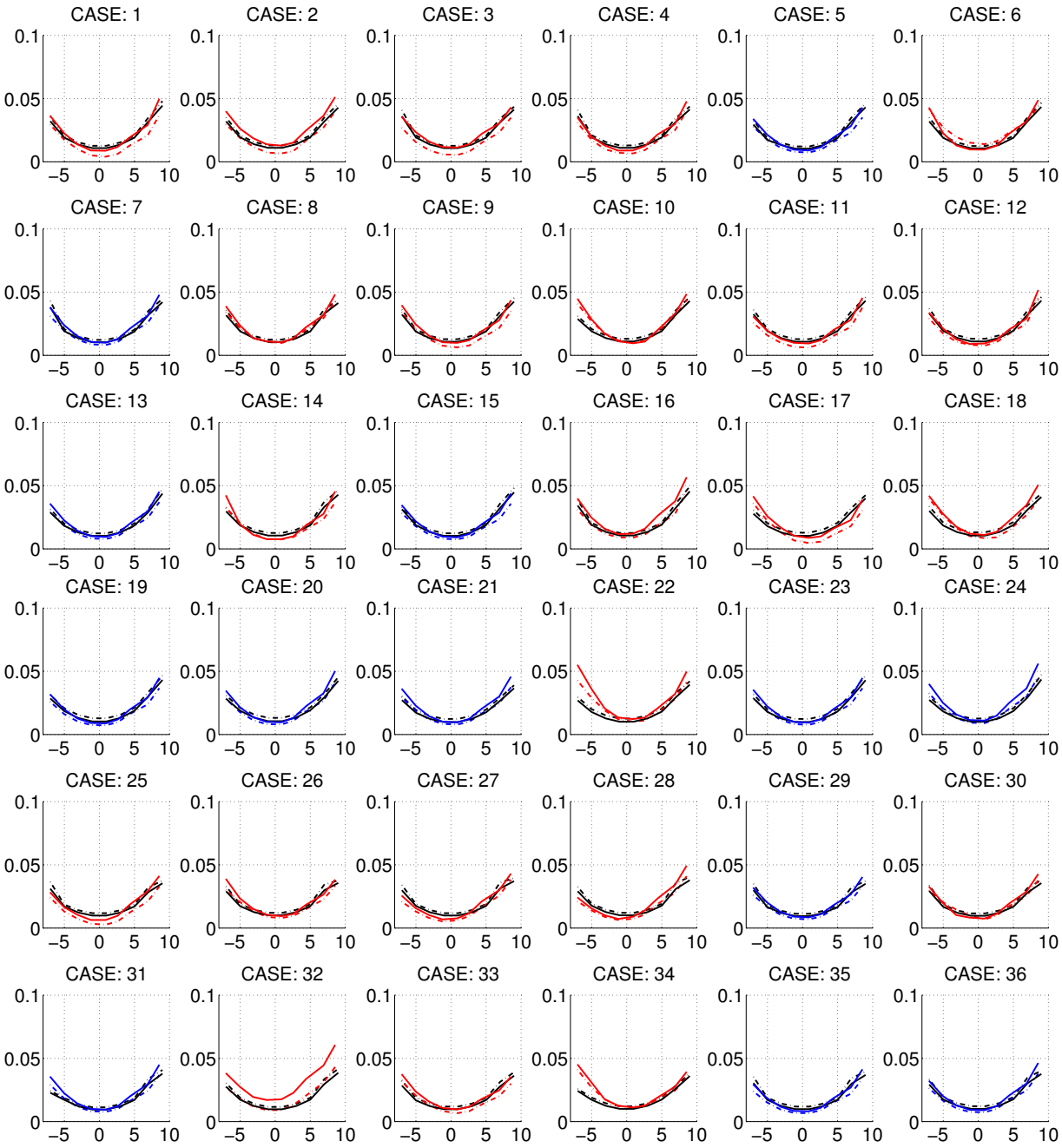

FIGURE B.7: Drag coefficient curves versus α for both Reynolds numbers $4.5 \cdot 10^5$ (dot-dashed lines) and $5.5 \cdot 10^5$ (solid lines), when the gappy-noisy-HOSVD method has been applied. Red lines represent the experimental database in the non-gappy configurations, blue lines represent the reconstructed C_D values in the gappy configurations using 12 modes and black lines represent numerical database. Half map of geometries: from case #1 to #36.

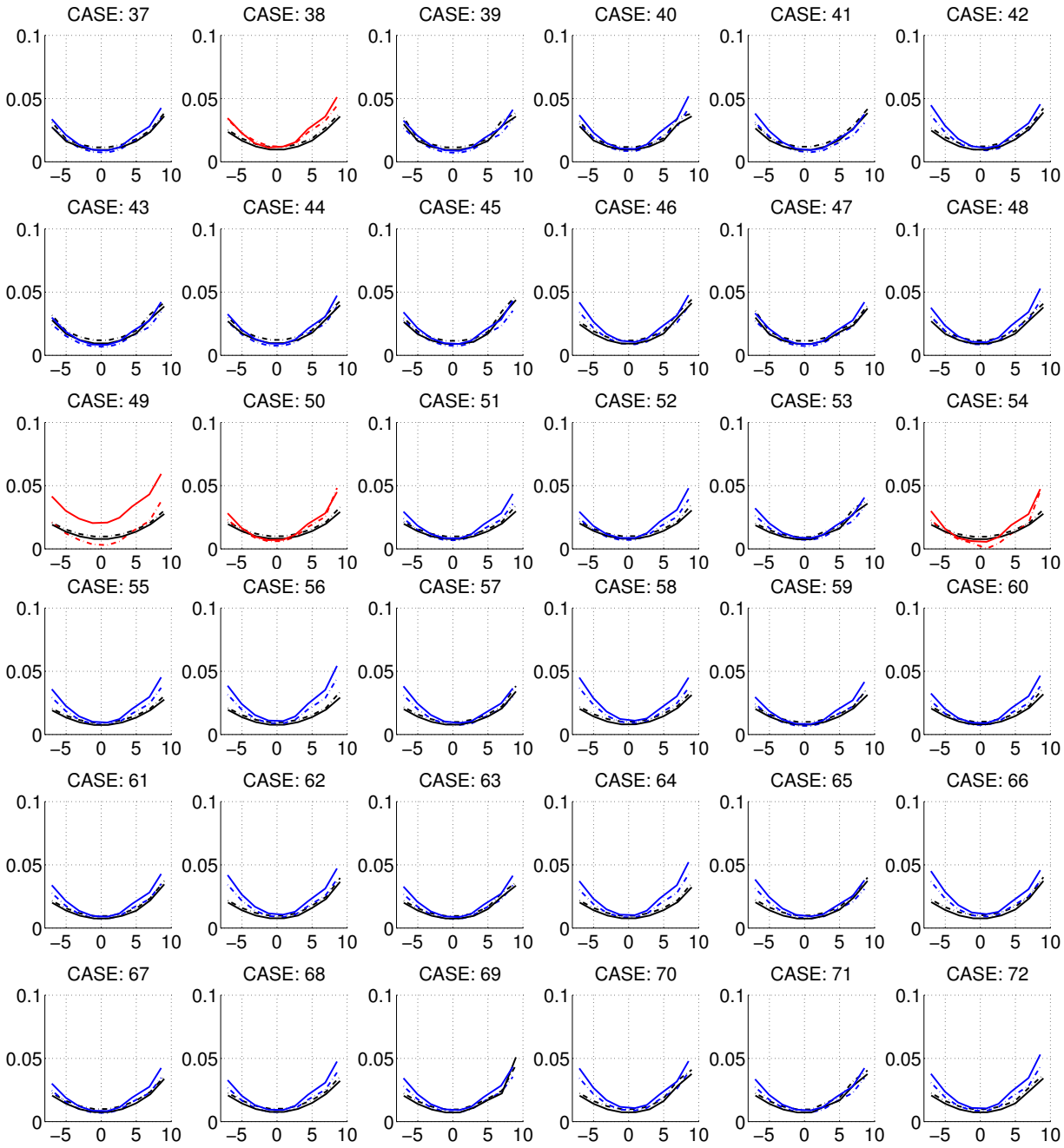

FIGURE B.8: Drag coefficient curves versus α for both Reynolds numbers $4.5 \cdot 10^5$ (dot-dashed lines) and $5.5 \cdot 10^5$ (solid lines), when the gappy-noisy-HOSVD method has been applied. Red lines represent the experimental database in the non-gappy configurations, blue lines represent the reconstructed C_D values in the gappy configurations using 12 modes and black lines represent numerical database. Half map of geometries: from case #37 to #72.

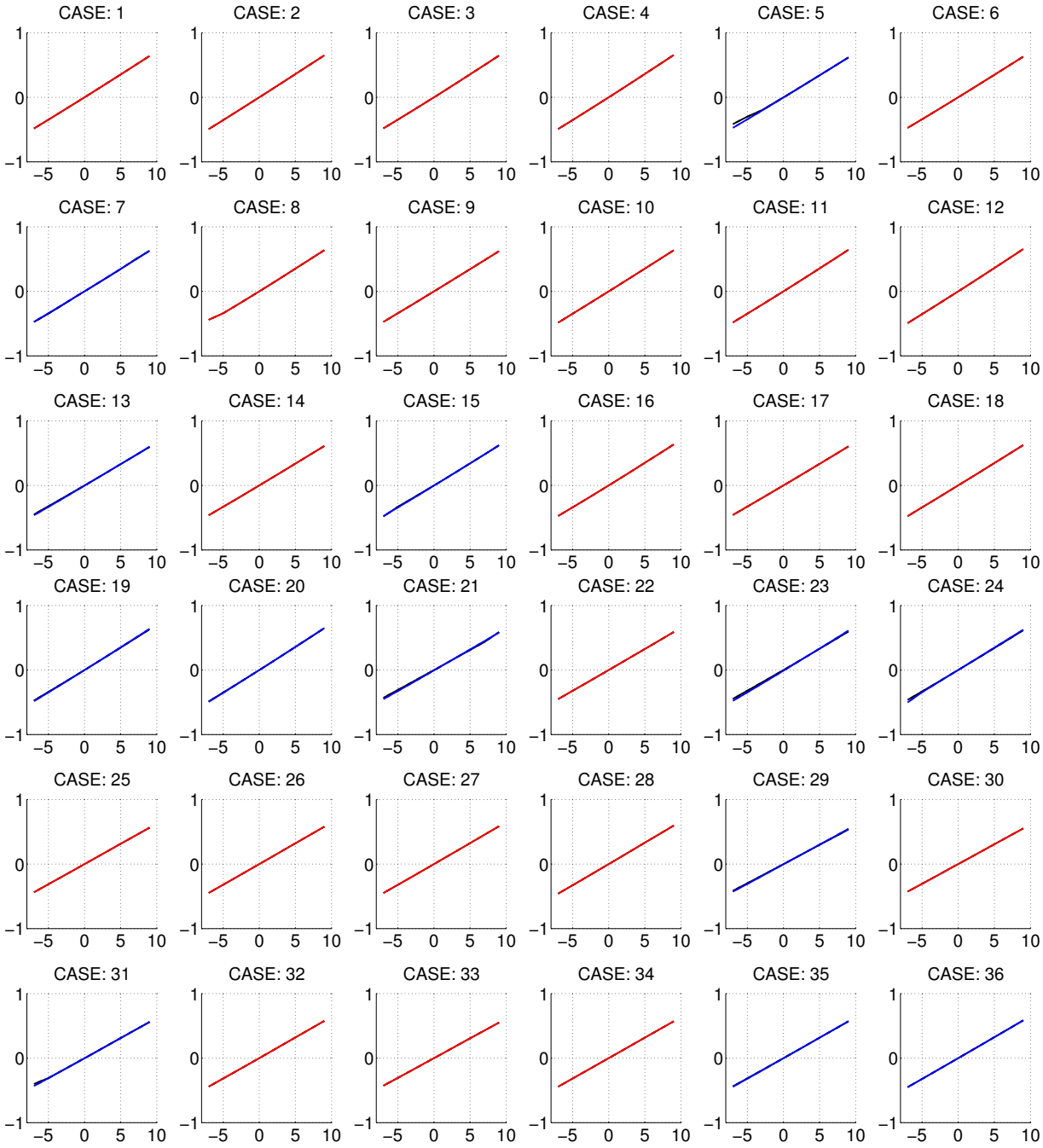

FIGURE B.9: Counterpart of Figure B.5 for the numerical database when same gappyness of the experimental database is present. The gappy region is reconstructed using 15 modes.

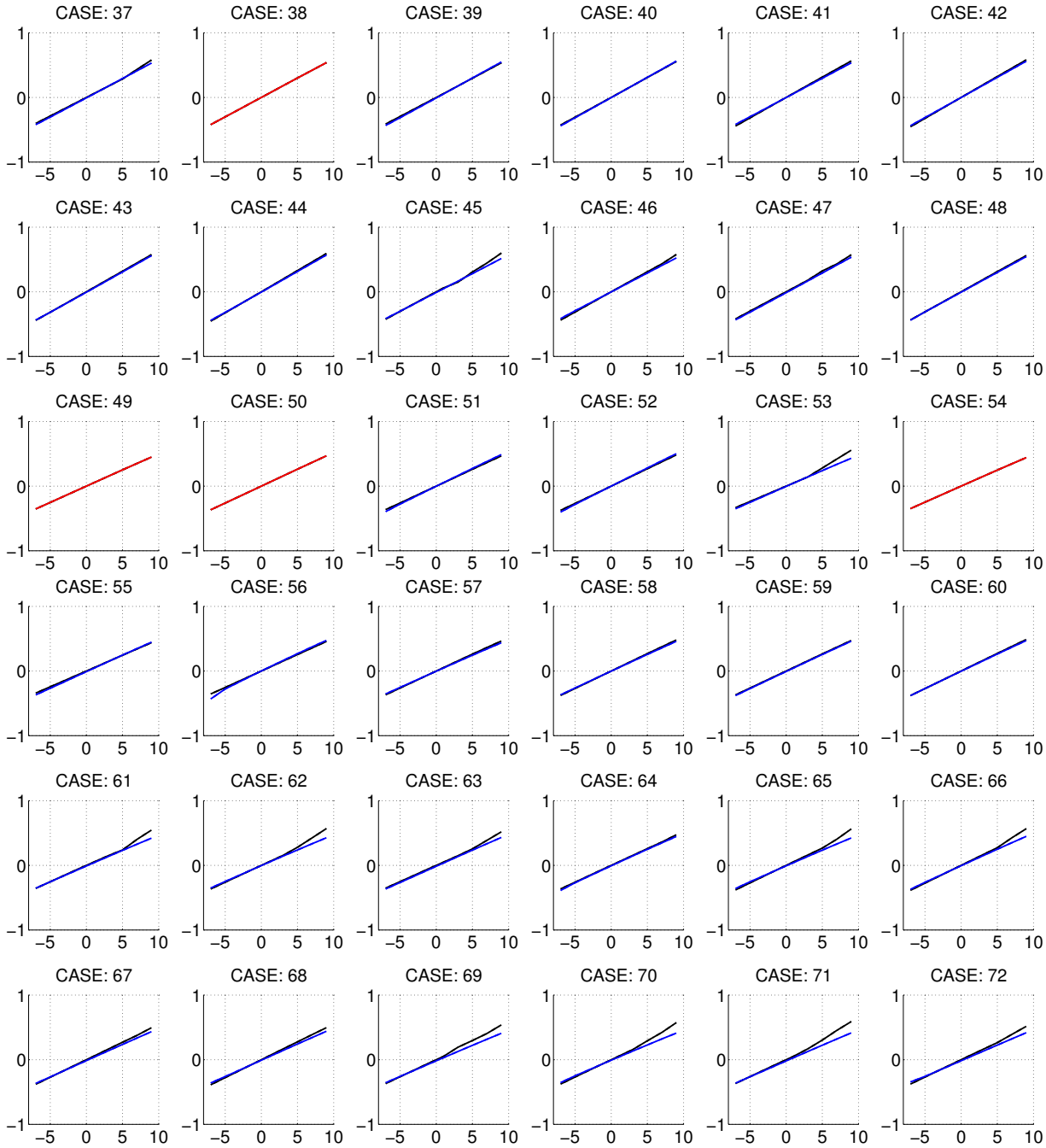

FIGURE B.10: Counterpart of Figure B.6 for the numerical database when same gappyness of the experimental database is present. The gappy region is reconstructed using 15 modes.

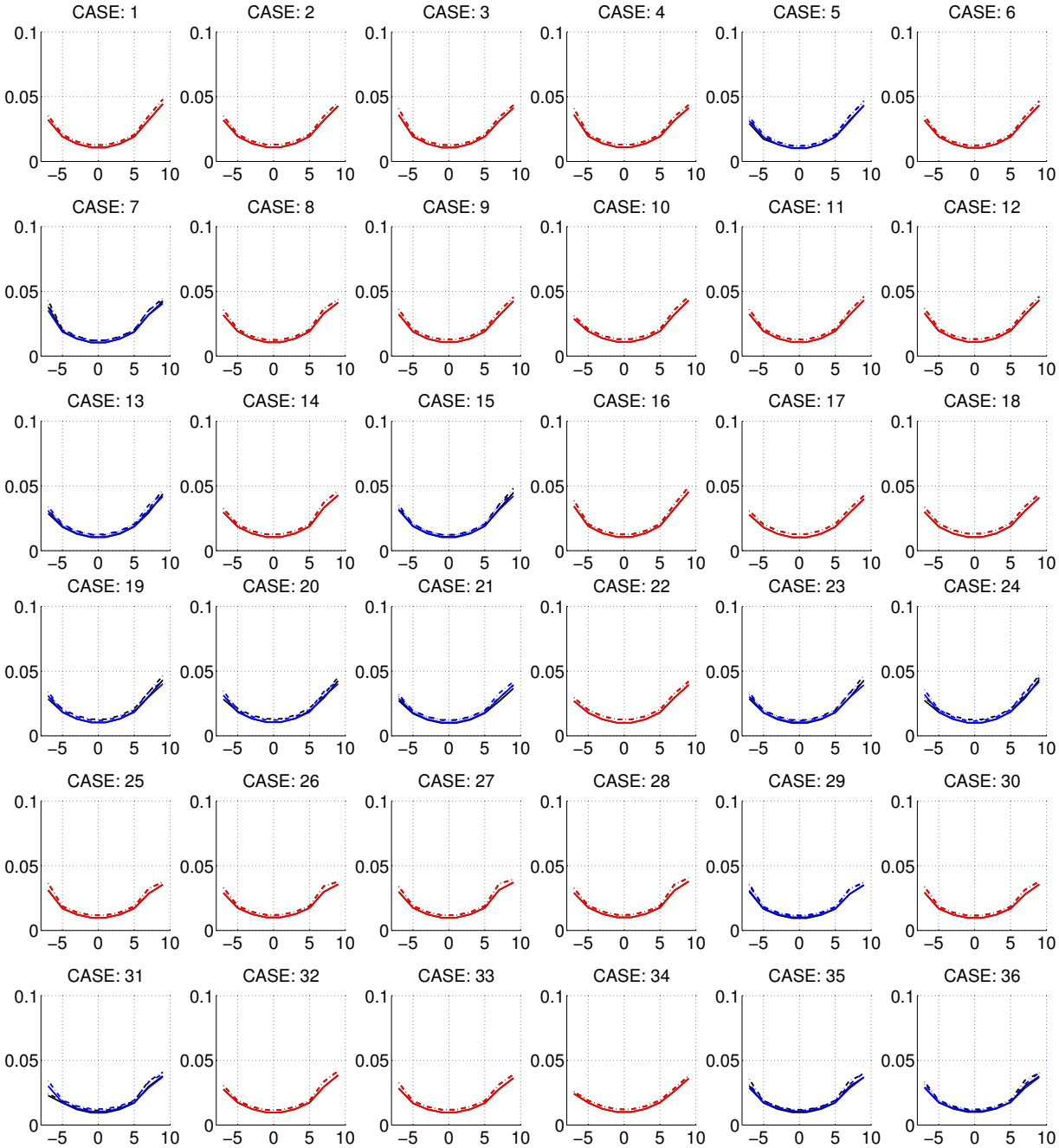

FIGURE B.11: Counterpart of Figure B.7 for the numerical database when same gappyness of the experimental database is present. The gappy region is reconstructed using 15 modes.

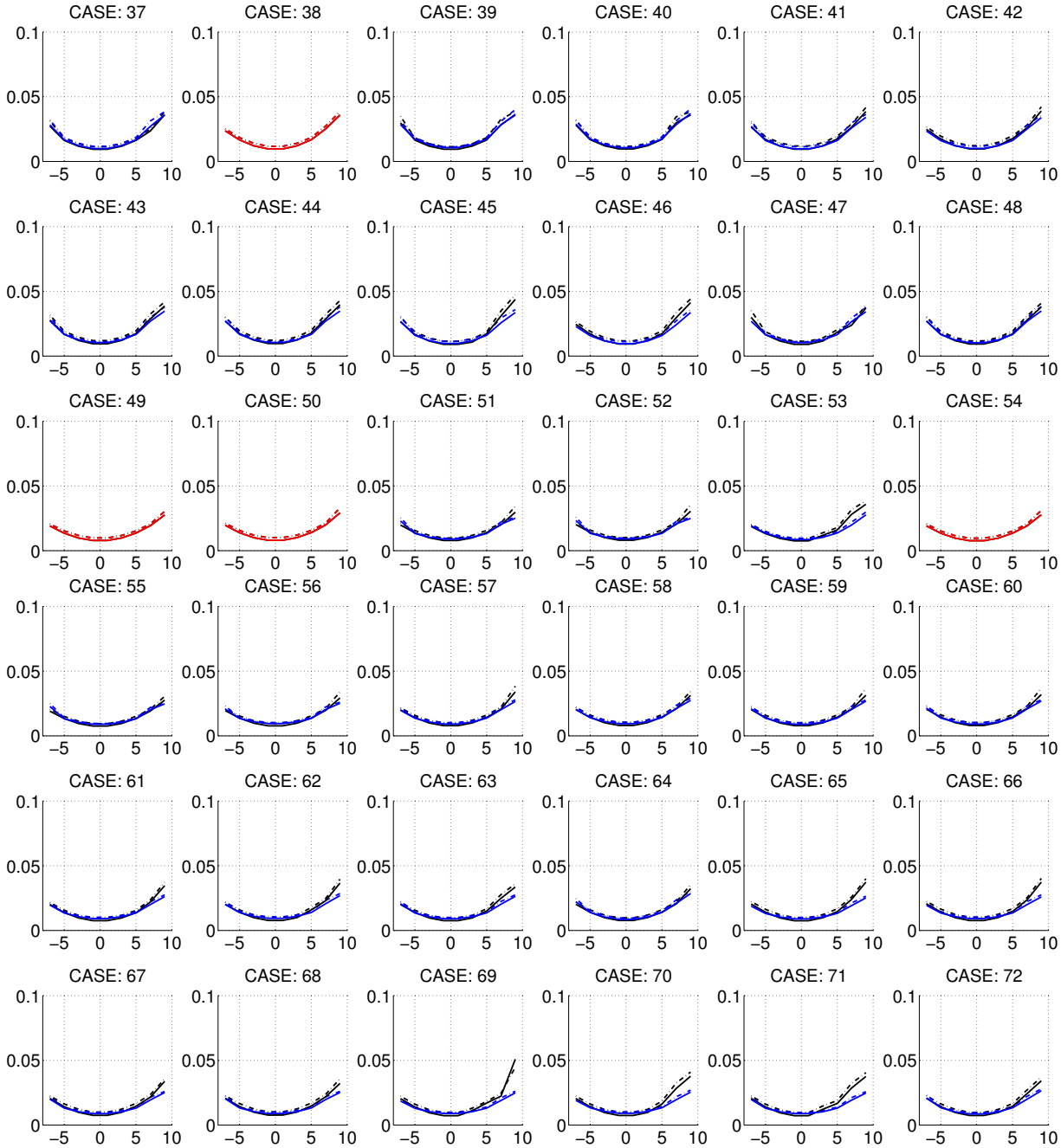

FIGURE B.12: Counterpart of Figure B.8 for the numerical database when same gappyness of the experimental database is present. The gappy region is reconstructed using 15 modes.

Bibliography

[1] Bellman, R., *Dynamic Programming*, Princeton University Press, Princeton, NJ, USA, 1st ed., 1957.

[2] Acar, E. and Yener, B., "Unsupervised Multiway Data Analysis: A Literature Survey," *IEEE Trans. Knowl. Data Eng.*, Vol. 21, No. 1, 2009, pp. 6 – 20.

[3] Tucker, L. R., *The extension of factor analysis to three-dimensional matrices*, Holt, Rinehart and Winston, New York, 1964, pp. 110 – 127.

[4] Tucker, L. R., "Some mathematical notes on three-mode factor analysis," *Psychometrika*, Vol. 31, No. 3, 1966, pp. 279–311.

[5] Kroonenberg, P. M., *Applied multiway data analysis*, Wiley series in probability and statistics, John Wiley and Sons, Inc., 2008.

[6] Gunes, H., Sirisup, S., and Karniadakis, G. E., "Gappy data: To Krig or not to Krig?" *J. Comput. Physics*, Vol. 212, 2006, pp. 358–382.

[7] Grinberg, L., Yakhot, A., and Karniadakis, G. E., "Analyzing Transient Turbulence in a Stenosed Carotid Artery by Proper Orthogonal Decomposition," *Annals of Biomedical Engineering*, Vol. 37, No. 11, 2009, pp. 2200 – 2217.

[8] Bui-Thanh, T., *Proper Orthogonal Decomposition Extensions and Their Applications in Steady Aerodynamics*, Master's thesis, Singapore-MIT Alliance, 2003.

[9] Bui-Thanh, T., Damodaran, M., and Willcox, K., "Aerodynamic Data Reconstruction and Inverse Design Using Proper Orthogonal Decomposition," *AIAA Journal*, Vol. 42, 2004, pp. 1505 – 1516.

[10] Lorente, L. S., Vega, J. M., and Velazquez, A., "Efficient computation of the POD manifold containing the information required to generate a multi-parameter aerodynamic database," *Aerospace Science and Technology*, Vol. 25, No. 1, 2013, pp. 152 – 160.

[11] Legresley, P. A. and Alonso, J. J., "Investigation of Non-Linear Projection for POD Based Reduced Order Models for Aerodynamics," *AAIA*, Vol. 2001-0926, 2001.

[12] Alonso, D., Velazquez, A., and Vega, J. M., "Robust reduced order modeling of heat transfer in a back step flow," *International Journal of Heat and Mass Transfer*, Vol. 52, No. 5 - 6, 2009, pp. 1149 – 1157.

[13] Alonso, D., Velazquez, A., and Vega, J. M., "A method to generate computationally efficient reduced order models," *Comput. Methods Appl. Mech. Engrg.*, Vol. 198, No. 33-36, 2009, pp. 2683 – 2691.

[14] Alonso, D., Vega, J. M., Velazquez, A., and de Pablo, V., "Reduced order modeling of three dimensional external aerdodynamic flows," *J. Aerospace Engrg.*, Vol. 25, No. 4, 2012, pp. 588 – 599.

[15] Ma, X. and Karniadakis, G. E., "A low-dimensional model for simulating three-dimensional cylinder flow," *J. Fluid Mech.*, Vol. 458, 2002, pp. 181 – 190.

[16] Braconnier, T., Ferrier, M., Jouhaud, J. C., Montagnac, M., and Sagaut, P., "Towards an adaptive POD / SVD surrogate model for aeronautic design," *Computers and Fluids*, Vol. 40, No. 1, 2011, pp. 195 – 209.

[17] Miled, Z. B., Li, H., Bukhres, O. A., Bem, M., Jones, R., and Oppelt, R. J., "Data Compression in a Pharmaceutical Drug Candidate Database," *Informatica*, Vol. 27, No. 2, 2003, pp. 213 – 224.

[18] del Castillo-Negrete, D., Hirshman, S. P., Spong, D. A., and D'Azevedo, E. F., "Compression of magnetohydrodynamic simulation data using singular value decomposition." *J. Comput. Physics*, Vol. 222, 2007, pp. 265–286.

[19] Sirovich, L. and Kirby, M., "Low-dimensional procedure for the characterization of human faces," *J. Opt. Soc. Am. A*, Vol. 4, No. 3, 1987, pp. 519 – 524.

[20] de Lathauwer, L., de Moor, B., and Vandewalle, J., "A multilinear singular value decomposition," *SIAM J. Matrix Anal. Appl*, Vol. 21, 2000, pp. 1253–1278.

[21] de Lathauwer, L., de Moor, B., and Vandewalle, J., "On the Best Rank-1 and Rank-(R1,R2,...,RN) Approximation of Higher-Order Tensors," *SIAM J. Matrix Anal. Appl.*, Vol. 21, 2000, pp. 1324 – 1342.

[22] Kolda, T. G. and Bader, B. W., "Tensor Decompositions and Applications," *SIAM Review*, Vol. 51, 2009, pp. 455–500.

[23] Golub, G. H. and Van Loan, C. F., *Matrix computations*, Johns Hopkins University Press, 3rd ed., 1996.

[24] Lorente, L. S., Vega, J. M., and Velazquez, A., "Generation of Aerodynamics Databases Using High-Order Singular Value Decomposition," *Journal of Aircraft*, Vol. 45, 2008, pp. 1779 – 1788.

[25] Lorente, L. S., Vega, J. M., and Velazquez, A., "Compression of aerodynamic databases using high-order singular value decomposition," *Aerospace Science and Technology*, Vol. 14, 2010, pp. 168 – 177.

[26] Bergqvist, G. and Larsson, E., "The Higher-Order Singular Value Decomposition: Theory and an Application," *IEEE Signal Process. Mag.*, Vol. 27, No. 3, 2010, pp. 151 – 154.

[27] Liu, X., De Lathauwer, L., Janssens, F., and De Moor, B., *Hybrid Clustering of Multiple Information Sources via HOSVD*, Springer Berlin Heidelberg, 2010, pp. 337 – 345.

[28] Benito, N., Arias, J. R., Velazquez, A., and Vega, J. M., "Real time performance improvement of engineering control units via Higher Order Singular Value Decomposition: Application to a SI engine," *Control Engineering Practice*, Vol. 19, No. 11, 2011, pp. 1315–1327.

[29] de Lucas, S., Vega, J. M., and Velazquez, A., "Aeronautic conceptual design optimization method based on High-Order Singular Value Decomposition," *AIAA Journal*, Vol. 49, No. 12, 2011, pp. 2713 – 2725.

[30] Yates, F., "The analysis of replicated experiments when the field results are incomplete," *Empire Journal of Experimental Agriculture*, Vol. 1, No. 2, 1933, pp. 129 – 142.

[31] Little, R. J. A. and Rubin, D. B., *Statistical Analysis with Missing Data*, John Wiley & Sons, Inc., 2nd ed., 2002.

[32] Everson, R. and Sirovich, L., "Karhunen-Loeve procedure for gappy data," *J. Opt. Soc. Am. A*, Vol. 12, 1995, pp. 1657–1664.

[33] Beckers, J. and Rixen, M., "EOF calculations and data filling from incomplete oceanographic datasets," *Journal of Atmospheric and Oceanic Technology*, Vol. 20, No. 12, 2003, pp. 1839–1856.

[34] Venturi, D. and Karniadakis, G., "Gappy data and reconstruction procedures for flow past a cylinder," *J. Fluid Mech.*, Vol. 519, 2004, pp. 315–336.

[35] Willcox, K., "Unsteady Flow Sensing and Estimation via the Gappy Proper Orthogonal Decomposition," *34th AIAA Fluid Dynamics Conference and Exhibit*, Portland, Oregon, 2004.

[36] Stein, M. L., *Interpolation of Spatial Data: Some Theory for Kriging*, Springer Series in Statistics, Springer, 1999.

[37] Algazi, V. and Sakrison, D., "On the optimality of the Karhunen-Loeve Expansion (Corresp.)," *IEEE Trans. Inform. Theory*, Vol. 15, No. 2, 1969, pp. 319 – 321.

[38] Yakhot, A., Anor, T., and Karniadakis, G. E., "A Reconstruction Method for Gappy and Noisy Arterial Flow Data," *IEEE Trans. Med. Imaging*, Vol. 26, 2007, pp. 1681 – 1697.

[39] Murray, N. E. and Ukeiley, L. S., "An application of Gappy POD," *Exp Fluids*, Vol. 42, No. 1, 2006, pp. 79 – 91.

[40] Venturi, D., Wan, X., and Karniadakis, G., "Stochastic low-dimensional modelling of a random laminar wake past a circular cylinder," *Journal of Fluid Mechanics*, Vol. 606, 2008, pp. 339–367.

[41] Raben, S. G., Charonko, J. J., and Vlachos, P. P., "Adaptive gappy proper orthogonal decomposition for particle image velocimetry data reconstruction," *Measurement Science and Technology*, Vol. 23, No. 2, 2012, pp. 025303.

[42] Jarzabek, A., Moreno, A., Perales, J. M., and M., V. J., "Database error filtering via SVD-like methods," *Aerospace Science and Technology - Under revision*, 2016.

[43] Astrid, P., Weiland, S., Willcox, K., and Backx, T., "Missing Point Estimation in Models Described by Proper Orthogonal Decomposition," *Automatic Control, IEEE Transactions on*, Vol. 53, No. 10, 2008, pp. 2237–2251.

[44] Yildirim, B., C., C., and Karniadakis, G. E., "Efficient sensor placement for Ocean measurements using Low-dimensional concepts," *Ocean Modelling*, Vol. 27, 2009, pp. 160 – 173.

[45] Stewart, G. W., "On the Early History of the Singular Value Decomposition," *SIAM Review*, Vol. 35, No. 4, 1993, pp. 551 – 566.

[46] Eckart, C. and Young, G., "The approximation of one matrix by another of lower rank," *Psychometrika*, Vol. 1, No. 3, 1936, pp. 211 – 218.

[47] Pearson, K., "On lines and planes of closest fit to systems of points in space," *Philosophical Magazine*, Vol. 2, No. 11, 1901, pp. 559 – 572.

[48] Chatterjee, A., "An introduction to the proper orthogonal decomposition," *Current Science*, Vol. 78, No. 7, 2000, pp. 808 – 817.

[49] da Silva, V. and Lim, L.-H., "Tensor Rank and the ill-Posedness of the Best Low-Rank Approximation Problem," *SIAM J. Matrix Analysis Applications*, Vol. 30, 2008, pp. 1084–1127.

[50] Roache, P., "Perspective: A Method for Uniform Reporting of Grid Refinement Studies," *Journal of Fluids Engineering*, Vol. 116, No. 3, 1994, pp. 405 – 413.

[51] Spalart, P. and Allmaras, S., *One equation turbulence model for aerodynamic flows*, No. AIAA-92-0439, 1992.

[52] Edwards, J. R. and Chandra, S., "Comparison of eddy viscosity-transport turbulence models for three-dimensional, shock-separated flowfields," *AIAA Journal*, Vol. 34, No. 4, 1996, pp. 756 – 763.

[53] Tannehill, J. C., Anderson, D. A., and Pletcher, R. H., *Computational fluid mechanics and heat transfer*, Taylor & Francis, 1997.

[54] Moukalled, F., Mangani, L., and Darwish, M., *The Finite Volume Method in Computational Fluid Dynamics: An Advanced Introduction with OpenFOAM® and Matlab*, Fluid Mechanics and Its Applications, Springer International Publishing, 2015.

[55] Abbott, I. H. and von, D. A. E., *Theory of wing sections*, Inc. Dover Publications, 2nd ed., 1959, pp. 47–63.

[56] Drela, M., "XFOIL: An Analysis and Design System for Low Reynolds Number Airfoils," *Conference on Low Reynolds Number Airfoil Aerodynamics*, Springer Berlin Heidelberg, 1989, pp. 1 – 12.